

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization International Bureau



(43) International Publication Date
12 February 2004 (12.02.2004)

PCT

(10) International Publication Number
WO 2004/013327 A1

(51) International Patent Classification⁷: C12N 15/10, (74) Agents: BECKER, Philippe et al.; Cabinet Becker et Associés, 35, rue des Mathurins, F-75008 Paris (FR).
C07K 14/00, C07H 21/00

(21) International Application Number:
PCT/EP2003/007765

(22) International Filing Date: 17 July 2003 (17.07.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
02291871.8 24 July 2002 (24.07.2002) EP

(71) Applicant (for all designated States except US): LIBRA-GEN [FR/FR]; 11, Boulevard Einstein, F-69100 Villeurbanne (FR).

(72) Inventors; and

(75) Inventors/Applicants (for US only): NALIN, Renaud [FR/FR]; 54, Avenue du Général De Gaulle, F-69300 Caluire (FR). PUJIC, Petar [HR/FR]; 4, rue Chantemerle, F-69360 Serezin Du Rhone (FR). TUPHILE, Karine [FR/FR]; 181, rue de l'Oratoire, F-69300 Caluire (FR). GILLET, Benjamin [FR/FR]; 46, rue de la Thibaudière, F-69007 Lyon (FR).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2004/013327 A1

(54) Title: METHOD FOR THE EXPRESSION OF UNKNOWN ENVIRONMENTAL DNA INTO ADAPTED HOST CELLS

(57) Abstract: The present invention relates to methods for the identification or cloning of polynucleotides encoding a selected phenotype, particularly from environmental DNA. In a specific embodiment, the method comprises (i) cloning environmental DNA fragments into Ecoli cloning vectors to produce a metagenomic library, (ii) identifying or selecting cloning vectors in said library which contain DNA fragments having a particular characteristic of interest, (iii) modifying the identified or selected cloning vectors into shuttle or expression vectors for transfer and integration in a selected host cell, (iv) transferring the modified cloning vectors into said selected host cell and (v) identifying or cloning the DNA fragments contained in said modified cloning vectors which encode said selected phenotype in said selected host cell.

Method for the expression of unknown environmental DNA into adapted host cellsIntroduction and Background

5 The present invention relates to methods and compositions for nucleic acid production, analysis and cloning. The present invention discloses tools and methods for the production and analysis of libraries of polynucleotides, particularly metagenomic libraries, which can be used to identify novel pathways, novel enzymes and novel metabolites of interest in various areas, including pharmaceutical, cosmetic, 10 agrochemical and/or food industry.

Drug discovery process is based on two main fields, namely combinatorial chemistry and natural products. Combinatorial chemistry has shown its ability to generate huge amounts of molecules, but with limited chemical diversity. At the opposite, natural 15 products have been the most predominant source of structural and molecular diversity. However, the exploitation of this diversity is strongly hampered by their limited access, complex identification and purification processes, as well as by their production.

Microorganisms are known to synthesize a large diversity of natural compounds which 20 are already widely used in therapeutic, agriculture, food and industrial areas. However, this promising approach to the identification of new natural compounds has always been considerably limited by the principal technological bolts of isolating and *in vitro* propagating the huge diversity of bacteria. Most microorganisms living in a natural, complex environment (soil, digestive tract, sea, etc...) have not been cultivated because 25 their optimal living conditions are either unknown or difficult to reproduce. Numbers of scientific publications relate this fact and it is now assumed that less than about 1% of the total bacterial diversity (when all environments are considered together) have been isolated and cultivated (Amann *et al*, 1995).

30 New approaches have been developed to try to overpass the critical step of isolation, and to access directly to the huge genetic potential established by the microbial adaptation

processes through their long evolution. These approaches are called "Metagenomic" because they address a plurality of genomes of a whole bacterial community, without any distinction (metagenome).

5 Metagenomics involve direct extraction of DNAs from environmental samples and their propagation and expression into a cultivated host cell, typically a bacteria. Metagenomic has been firstly developed for the identification of new bacterial phylum (Pace .1997). This use is based on the specific cloning of genes recognized for their interest as phylogenetic markers, such as 16S rDNA genes. Further developments of Metagenomics
10 relate to the detection and cloning of genes coding for proteins with environmental or industrial interest. These first two applications of metagenomic involve a first step of gene selection (generally using PCR) before cloning. In the case of protein production, the cloning vector used are preferentially also expression vectors, i.e., they contain regulatory sequences upstream of the cloning site causing expression of the cloned gene
15 in a given bacterial host strain.

More recent developments of metagenomic consider the total metagenome cloned without any selection and/or identification, to establish random "Metagenomic DNA libraries". This provides an access to the whole genetic potential of bacterial diversity without any "*a priori*" selection. Metagenomic DNA libraries are composed of hundreds
20 of thousands of clones which differ from each other by the environmental DNA fragments which have been cloned. In this respect, large DNA fragments have been cloned (more than 30 Kb), so as to (i) limit the number of clones which have to be analysed and (ii) to be able to recover whole biosynthetic pathways for the identification of new metabolites resulting from multi enzymatic synthesis. This last point is of
25 particular interest for bacterial metagenomic libraries since, most the biosynthetic pathways have been found to be naturally organised in a same cluster of DNA and even in the same operon in bacteria. Nevertheless, the heterologous expression of a whole biosynthetic pathways (large DNA fragment) needs a much more improved system than a simple expression vector to have a full and stable expression.

Except for the identification and characterisation of bacterial community at the phylogenetic or diversity levels, metagenomic libraries produced in the prior art are gene expression libraries, i.e., the environmental DNA fragments are cloned downstream of a functional promoter, to allow their expression and analysis. In this regard, WO99/45154 5 and WO96/34112 relate to combinatorial gene expression libraries which comprise a pool of expression constructs where each expression construct contains DNA which is operably associated with one or more regulatory regions that drive expression of genes in an appropriate organisms. Furthermore, the expression constructs used in these methods have a very limited and invariable host range. Similarly, WO 01/40497 relates to the 10 construction and use of expression vectors which can be transferred in one chosen expression bacterial host of the *Streptomyces* genus. All these approaches are, however, very limited since they require the presence of expression signals and confer invariable or very limited host range capabilities. Furthermore, most (if not all) metagenomic DNA libraries have been established in *E. coli* which is the most efficient cloning system. 15 However, most environmental DNA are not expressed or functionally active in *E. coli*. In particular, functional analysis in *E. coli* of genes cloned from G+C rich organisms, such as *Actinomyces*, could be limited by the lack of adequate transcription and translation system. Also, posttranslational modification system in *E. coli* is not operative on heterologous proteins from *Actinomycetes* and some specific substrates for proteins 20 activity are not present in *E. coli*.

The stable maintenance of large foreign DNA fragments (> 10Kb) into a selected host cell is one of the key points for academic research or applied industrial purposes. Usually, the vector carrying the foreign DNA is maintained by cultivating the host cells 25 in a medium with a vector-specific selective pressure (resistance to an antibiotic for example). However, when large foreign DNA fragments are cloned and/or expressed, their propagation and/or expression require energy, which is not allocated for cell growth anymore. As a consequence of this new resource allocation (nutrients/energy), it is not unusual to have a genetic rearrangement of the foreign DNA (deletion, modification 30 etc...) as a recombinant cell reaction. This results in the modification of the foreign genetic information and in the loss of DNA functionality. This can be observed without

any loss of the selective pressure carried by the vector. As a result, the recombinant clone is no more exploitable for genetic or functional analysis.

Thus, the exploitation of the huge potential of metagenomics for the discovery of new natural compounds, pathways or genes cannot be achieved with currently existing methods. Alternative technologies and processes must be developed, to allow stable maintenance and propagation of large foreign DNAs into host cells for production of efficient libraries and functional screening in a large variety of host cell species, including *Bacillus* or *Streptomyces*, to take full account of the huge diversity of the environmental DNAs.

Summary of the Invention

The present invention discloses improved tools and methods for the production and analysis of libraries of polynucleotides, particularly metagenomic libraries, which can be used to identify and produce novel pathways, novel enzymes and novel metabolites of interest.

More particularly, the invention now proposes to keep the advantage of high efficient cloning in *E. coli* and to modify the properties of metagenomic libraries, to allow genetic and functional analyses of particular selected clones in any appropriate system, thereby making possible the stable maintenance and propagation, the analysis and/or the expression of the huge diversity of metagenomic libraries. According to the invention, polynucleotide libraries can be produced in any convenient cloning system, such as *E. coli*, and then modified, depending on the desired selection or screening system, to adapt host range and/or properties of the library (or a portion thereof).

A particular object of this invention resides more specifically in a method of analysing a library of polynucleotides, said polynucleotides being contained in cloning vectors having a particular host range, the method comprising (i) selecting cloning vectors in the library which contain a polynucleotide having a particular characteristic, (ii) modifying

said selected cloning vectors to allow a transfer of said vectors into a selected host cell and integration of the polynucleotide contained in said vectors into the genome of the selected host cell, and (iii) analysing the polynucleotides contained in said modified vectors upon transfer of said modified vectors into said selected host cell.

5

An other object of this invention is a library of polynucleotides, wherein said library comprises a plurality of environmental DNA fragments cloned into cloning vectors, wherein said environmental DNA fragments contain a common molecular characteristic and wherein said cloning vectors are *E. coli* cloning vectors comprising a target polynucleotide construct allowing (i) transfer of the environmental DNA into a selected host cell distinct from *E. coli*, (ii) integration of the environmental DNA into the genome of a selected host cell, and (iii) stable maintenance and propagation of the environmental DNA into the selected host cell.

10 15 A further object of this invention is a method of producing modified libraries of polynucleotides, the method comprising selecting a sub-population of clones in a first library, based on the presence or absence of a characteristic of interest, and modifying the properties of said selected clones to allow their functional analysis or expression.

20 The modification in the library or cloning vector is typically obtained by targeted insertion of a polynucleotide construct, preferably using transposable elements, either in vitro or in vivo.

25 The integration into the genome of the selected host cell is typically obtained by site specific integration or by homologous or heterologous DNA/DNA recombination.

The invention is particularly suited for producing and analysing genetic diversity (metagenomic libraries), to identify new genes and isolate new metabolites, drugs, enzymes, antibiotics, etc.

Legend to the Figures

Figure 1 : Map of pP1 vector carrying transposable construct Tn<Apra> fig.1a. Excised transposable construct is 1132 bps in size. It contains two mosaic ends (ME) and a gene 5 conferring resistance to apramycin (Apra) fig.1b.

Figure 2 : pPL1 Vector (fig. 2a) that carries the conjugative transposable construct Tn<Apra-oriT> (fig. 2b). The nucleotide sequence of the transposable construct contains an origin of transfer (oriT) and a gene conferring resistance to apramycin (Apra). 10 Direction of DNA transfer at oriT is shown by an arrow.

Figure 3 : Transposable construct Tn<Apra-oriT-att-int> (fig. 3a) on vector pPAOI6 (fig. 3b). The transposable construct contains Φ C31 integrase gene and attachment DNA sequence for site specific integration, origin of transfer and gene for selection. The 15 orientation of genes and direction of DNA transfer are marked by arrows.

Figure 4 : This figure shows any target DNA suitable for insertion of transposable construct (fig. 4a). Insertion of conjugative and site specific integrative transposable construct Tn<Apra-oriT-int> is shown on fig. 4 b and c. Insertion of transposable construct into selective gene marker carried on original vector is shown on fig. 4b. In this 20 event, the cloned insert is intact and can be transferred to heterologous host. Insertion of transposable construct into cloned DNA insert results in gene inactivation (fig. 4c).

Figure 5 : Complete annotated DNA sequences of fosmid clones FS3-124 (a, SEQ ID 25 NO: 1) and FS3-135 (b, SEQ ID NO: 2).

Figure 6 : Morphological differences between *Streptomyces* transconjugant (assay). Conjugations have been performed with FS3-124 modified with transposable construct pPAOI6; (control) conjugation have been performed with pPAOI6.

Figure 7: Schematic Map of pPSB vector (top) and transposable elements (bottom). The transposable element has 720 bp DNA of the *amyE* gene from *B. subtilis* in addition to *att-int-oriT-Apra*.

5 Figure 8: Plasmid pPSBery (top) and transposable element (bottom) carrying selective marker *ery* AM, for resistance to erythromycine, and part of *amy E* gene for homologus recombination in *B. subtilis*.

10 Figure 9: Integrase Φ C31 was deleted from pPSBery plasmid. Resulting plasmid is pPSBery-DI (top). Transposable element contains *Apra* and *ery* AM genes for selection, oriT origin of transfer and a part of *amyE* gene for integration in to *amyE* locus of *B. subtilis* chromosome (bottom).

15 Figure 10: Map of pTn5-7 AOI plasmid. Transposable element has ends of tn5 (ME) and tn7 (T7 R, T7 L) transposons.

Detailed Description of the Invention

20 The invention provides novel strategies, methods and products for generating and analysing combinatorial gene libraries. As indicated above, the invention discloses, particularly, methods of analysing libraries of polynucleotides, said polynucleotides being contained in cloning vectors having a particular host range, the methods comprising (i) selecting cloning vectors in the library which contain a polynucleotide having a particular characteristic, (ii) modifying said selected cloning vectors to allow a 25 transfer of said vectors and/or expression of the polynucleotide which they contain into a selected host cell, and (iii) analysing the polynucleotides contained in said modified vectors upon transfer of said modified vectors into said selected host cell, such as by genetic, biochemical, chemical or phenotypical approaches.

30 In a most preferred embodiment, the methods allow stable transfer and propagation of large environmental nucleic acids in a selected host following initial selection. Such

methods comprise (i) selecting cloning vectors in a library which contain a polynucleotide having a particular characteristic, (ii) modifying said selected cloning vectors to allow a transfer of said vectors and integration of the polynucleotide which they contain into a selected host cell genome, and (iii) analysing the polynucleotides contained in said modified vectors upon transfer of said modified vectors into said selected host cell, such as by genetic, biochemical, chemical or phenotypical approaches.

Library of polynucleotides

10 The term "library of polynucleotides" designates a complex composition comprising a plurality of polynucleotides, of various origins and structure. Typically, the library comprises a plurality of unknown polynucleotides, i.e., of polynucleotides whose sequence and/or source and/or activity is not known or characterized. In addition to such unknown (or uncharacterized) polynucleotides, the library may further include known sequences or polynucleotides. Typically, the library comprises more than 20 distinct polynucleotides, more preferably at least 50, typically at least 100, 500 or 1000. The complexity of the libraries may vary. In particular, libraries may contain more than 5000, 10 000 or 100 000 polynucleotides, of various origin, source, size, etc. Furthermore, the polynucleotides are generally cloned into cloning vectors, allowing their maintenance and propagation in suitable host cells, typically in *E. coli*. The polynucleotides in the library may be in the form of a mixture or separated from each other, in all or in part. It should be understood that some or each polynucleotide in the library may be present in various copy numbers.

15 20 25 The polynucleotides in the libraries are more preferably obtained or cloned from complex sources of nucleic acids, most preferably from environmental samples. Such libraries are also termed "metagenomic libraries" since they contain nucleic acids derived from whole genomes of mixed populations of microorganisms.

30 The term environmental sample designates, broadly, any sample containing (a plurality of) uncharacterized (micro)organisms, particularly uncultivated (or non-cultivable)

microorganisms. The sample may be obtained or derived from specific organisms, natural environments or from artificial or specifically created environments (e.g., industrial effluents, etc). An uncultivated (or non-cultivable) microorganism is a microorganism that has not been purposely cultured and expanded in isolated form. The 5 sample may be obtained or derived from soil, water, mud, vegetal extract, wood, biological material, marine or estuarine sediment, industrial effluents, gas, mineral extracts, sand, natural excrements, meteorites etc. The sample may be collected from various regions or conditions, such as tropical regions, deserts, volcanic regions, forests, farms, industrial areas, household, etc.

10

Environmental samples usually contain various species of (uncharacterized, uncultivated) microorganisms, such as terrestrial microorganisms, marine microorganisms, salt water microorganisms, freshwater microorganisms, etc. Species of such environmental microorganisms include autotrophe or heterotrophe organisms, 15 eubacteria, archaebacteria, algae, protozoa, fungi, viruses, phages, parasites, etc. The microorganisms may include extremophile organisms, such as thermophiles, psychrophiles, psychrotrophes, acidophiles, halophiles, etc. More specific examples of environmental bacteria includes actinomycetes, eubacteriaes and mycobacteriaes, examples of fungi include phycomycetes, ascomycetes and basidiomycetes, etc. Other 20 organisms include yeasts (saccharomyces, kluyveromyces, etc.) plant cells (algae, lichens, etc), corals, etc. for instance. The sample may comprise various species of such donor (uncultivated) microorganisms, as well as various amounts thereof. The environmental sample may contain, in addition, known and/or cultivable microorganisms (e.g., prokaryotic or eukaryotic); as well as nucleic acids and organic 25 materials. The sample may also contain different animal cells: mammalian cells; insect cells, etc (arising from larvae, feces, etc.).

It should be understood that the present invention is not limited to any specific type of sample or environmental microorganism, but can be used to produce diversity, create 30 nucleic acid libraries, etc., from any environmental sample comprising uncultivated microorganisms. The sample may be wet, soluble, dry, in the form of a suspension,

paste, powder, solid, etc. Preferably, the sample is dry or in solid or semi-solid state (e.g., paste, powder, mud, gel, etc.). The sample may be treated prior to nucleic acid extraction, for instance by washings, filtrating, centrifuging, diluting, drying, etc.

5 The term "environmental DNA" designates any DNA fragment or collection obtained from an environmental sample. Nucleic acids may be extracted / isolated from the sample according to various techniques, such as those described in WO01/81357, in WO01/40497, in Handelsman *et al.* (Chemistry & Biology 5(10), 1998, R245), Rondon *et al.* (Tibtech 17, 1999, 403 ; Applied and Environm. Microbiol. 66, 2000, 2541), Miller *et al* (Applied and Environm. Microbiol. 65, 1999, 4715) or Frostegard *et al.* (Applied and Environm. Microbiol., 65, 1999, 5409).

10

15 In a particular embodiment of the above method, the library comprises a plurality of environmental DNA fragments. The library may also comprise other types of nucleic acids, such as environmental RNAs, for instance.

20 The polynucleotides have a size which is typically comprised between 10 and 100 kb, more preferably between 20 and 80 kb, typically between 30 and 80 kb. Although not mandatory, it is preferred that the polynucleotide fragments in the library all have similar size, to produce homogenous libraries.

Cloning Vectors

25 As indicated, the polynucleotides are contained or cloned into cloning vectors. These vectors may be of various types, including plasmids, cosmids, fosmids, episomes, artificial chromosomes, phages, viral vectors, etc. In most preferred embodiments, the cloning vectors are selected from plasmids, cosmids, phages (e.g., P1-derivatives) and BACs, even more preferably from cosmids, P1 derivatives and BACs. By using cosmids or P1 derivatives, it is possible to generate homogenous libraries, since these vectors essentially accommodate polynucleotides having a size of approximately 40 kb and 80 Kb, respectively. Furthermore, since the invention provides that the vectors are modified

30

after the initial cloning step, the cloning capacity of the vectors is maximized and inserts of as much as 40 and 80 kbs in length can be cloned into fosmids, BAC and P1 derivatives.

5 As indicated the cloning vector has a particular host range, i.e., the ability to replicate in a particular type of host cell. Typically, the host is a bacteria, more preferably an *E. coli* strain. Indeed, *E. coli* is so far the most convenient host cell for performing recombinant technologies. The advantage of the present invention is that the starting library can be produced in any suitable host system of choice, since the properties of the libraries will
10 be adapted later during the process.

Cloning vectors generally comprise the polynucleotide insert and genetic elements necessary and sufficient for maintenance into a competent host cell. They typically contain, in addition to the polynucleotide insert, an origin of replication functional in a
15 selected host cell as well as a marker gene for selection and screening. The cloning vector may comprise additional elements, such as promoter regions, for instance. Although cloning vectors may replicate in several different host cells, they are usually adapted to a particular host cell type and not suitable or efficient for replication or maintenance in other cell types.

20 In a preferred embodiment, the cloning vectors of the library are *E. coli* cloning vectors, preferably cosmids, BAC or P1 derived vectors. *E. coli* cloning vectors may carry an origin of replication derived from naturally-occurring plasmids, such as ColE1, pACYC and p15A, for instance. Many *E. coli* cloning vectors are commercially available and/or
25 can be constructed using available regulatory sequences.

Screening of the cloning vectors

In step i) of the method, a first selection or screen is performed on the polynucleotide
30 library. The screen is performed so as to identify or select clones having (or lacking) a particular, common characteristic. The selection may be carried out according to various

techniques, such as molecular screening, protein expression, functional screening, etc. A preferred selection is performed by molecular screening. Molecular screening designates any method of identification of molecular or structural characteristics in a polynucleotide sequence. This can be made by a variety of techniques which are known per se, such as 5 hybridisation, amplification, sequencing, etc. Preferably, molecular screening comprises the selection of clones in a library which contain, in their sequence, a particular sequence or region or motif, said sequence or region or motif being characteristic of a particular type of activity or gene (enzyme, biosynthetic pathways, etc.).

10 In a first variant, the selection is made by contacting the cloning vectors in the library with a particular nucleic acid probe (or set of probes) containing a sequence which is characteristic of a selected activity or function (a consensus sequence, a particular motif, etc.). The cloning vectors in the library which hybridise to the probe (or set of probes) are then selected.

15 In a second variant, the selection is made by contacting the cloning vectors in the library with a particular pair of nucleic acid primers specific for a sequence which is characteristic of a selected activity or function (a consensus sequence, a particular motif, etc.), and a PCR amplification reaction is performed. The cloning vectors in the library 20 which lead to a positive amplification product are then selected.

In this regard, the present application provides new primers designed in conserved motives of the β -keto acyl synthase gene, which are particularly useful for screening polynucleotides containing putative polyketide synthase (PKS) genes or domains. These 25 primers have the following degenerated sequence :

Sense primer : 5'- GGSCCSKCSSTSDCSRTSGAYACSGC -3' (SEQ ID NO: 3)

Antisense primer : 5'- GCBBSRYYTCDATGGRTCSCC -3' (SEQ ID NO: 4)

wherein :

R is A or G

30 S is G or C

Y is C or T

K is G or T

D is A or G or T, and

B is C or G or T.

5 A particular object of this invention is a polynucleotide primer having one of the above sequences, typically a mixture of different polynucleotide primers having a sequence corresponding to one of the above degenerated sequences. A particular object of this invention also resides in a pair of primers each having one of the above sequences.

10 Once particular clones have been selected, the analysis of their polynucleotides needs to be confirmed and/or validated, and/or their polynucleotides can be used to study their function and/or produce novel compounds or metabolites.

15 The invention now enables such further analysis and uses, by allowing a modification of the cloning vectors that is specific and adaptable by the skilled person, depending on the activity which is sought. In particular, it is possible to confer properties such as specific expression or a novel, specific host range to the selected cloning vectors, to assess their activity, as disclosed below.

20 Modification of the cloning vectors

After high efficiency cloning using most convenient cloning vectors such as BACs or cosmids propagated into *E. coli*, and after the identification, selection and/or characterisation of cloned DNA fragments, the invention now allows to modify 25 specifically the cloning vectors to transfer, integrate into the genome, maintain, express and/or over-express the selected polynucleotides into any selected host expression system, which is suitable to assess the selected activity or property. Such selected hosts may be native or heterologous host cells, and include, but are not limited to, for example *Streptomyces*, *Nocardia*, *Bacillus*, fungi, yeasts, etc.

The selected cloning vectors of the library may be modified according to various techniques. The modification is typically a genetic modification, comprising the introduction of particular genetic sequences into the structure of the cloning vector, in addition to or in replacement of sequences contained in said vector. It is highly preferred 5 to use specific or targeted (or oriented) techniques to improve the efficacy of the method. By "specific" is meant that the modification occurs at a pre-determined location in the cloning vector, through site-specific mechanisms. By "targeted" is meant that the modification occurs in a controlled way, so as not to alter the polynucleotide insert contained in the vector in a non-desirable way.

10

In a preferred embodiment, the selected vectors are modified by insertion, into the vector, of a target polynucleotide construct which contains genetic elements conferring the selected property(ies) to the cloning vector.

15

The target polynucleotide construct typically comprises the genetic elements necessary to transfer, propagate, integrate into the genome, maintain, express or overexpress the cloned polynucleotide into a chosen (bacterial) host expression system. Said genetic elements may include particular origin(s) of replication, particular origin(s) of transfer, particular integrase(s), transcriptional promoter(s) or silencer(s), either alone or in 20 combination(s).

In a first, preferred variant, the target polynucleotide construct comprises a genetic element allowing transfer of the vector into a selected host cell.

25

Natural DNA transfer mechanisms between donor and recipient strains is known under the term conjugation or conjugative transfer. Conjugative transfer can occur between different strains of the same species as well as between strains of different species. Many naturally occurring plasmids carry so called *tra* genes, which are involved in and mediate conjugative transfer. The DNA transfer starts at specific DNA structures, known as an 30 origin of transfer or "ori T". The presence of such an *oriT* in a vector allows said vector to be transferred into a desired host cell.

In a particular, preferred embodiment, the target polynucleotide construct comprises an origin of transfer functional in the selected host cell.

The structure of various oriT has been reported in the art (Guiney *et al.* 1983; Zechner *et al.* 2000). In a specific embodiment, the origin of transfer is selected (or derived) from RP4, pTiC58, F, RSF1010, ColE1 and R6K(α).

A specific example of an oriT which can be used in the present invention derives from plasmid RP4 and has or comprises all or a functional part of the following sequence 10 (SEQ ID NO: 5) :

gatctGTGATGTACTTCACCAGCTCCGCGAAGTCGCTCTTGTATTGGAGCGCATGGG
GACGTGCTTGGCAATCACGCGACCCCCCGGCCGTTTAGCGGCTAAAAAAAGTCAT
GGCTCTGCCCTCGGGCGGACCACGCCATCATGACCTTGCCAAGCTCGTCCTGCTTC
15 TCTTCGATCTCGCCAGCAGGGCGAGGATCGTGGCATACCGAACCGCGCCGTGCG
CGGGTCGTCGGTGAGCCAGAGTTTCAGCAGGCCGCCAGGCGGCCAGGTGCGCAT
TGATGCGGGCCAGCTCGCGACGTGCTCATAGTCCACGACGCCGTGATTGGTAGC
CCTGGCCGACGGCCAGCAGGTAGGCCACAGGCTCATGCCGGCCGCCGCCCTT
20 TCCTCAATCGCTCTCGTCTGGAAAGGCAGTACACCTTGATAGGTGGGCTGCC
TTCCTGGTTGGCTTGGTTCATCAGCCATCCGCTGCCCTCATCTGTTACGCCGGCGG
TAGCCGGCCAGCCTCGCAGAGCAGGATTCCCGTTGAGCACCAGGTGCGAATAA
GGGACAGTGAAGAAGGAACACCCGCTCGCGGGTGGGCCTACTTCACCTATCCTGCC
25 CGGCTGACGCCGTTGGATACACCAAGGAAAGTCTACACGAACCCCTTGGCAAAATC
CTGTATATCGTGCAGGGAAAGGATGGATATACCGAAAAAATCGCTATAATGACCCCG
AAGCAGGGTTATGCAGCGGAAAAGATCCGTCGGATCT

The term "functional part" designates any fragment or variants of the above sequence which retain the capacity to cause conjugative transfer. Such fragments typically comprise at least 80%, preferably at least 85% or 90% of the above sequence. Variants 30 may include one or several mutations, substitutions, deletions or additions of one or several bases.

In an other particular variant, the target polynucleotide construct comprises a genetic element allowing integration of the vector (or of the polynucleotide contained therein) 35 into the genome of the selected host cell.

A donor DNA is permanently or stably maintained and expressed in a selected recipient cell if it is integrated into the recipient cell's genome or if it contains elements that allow autonomous replication in said cell. In a most preferred embodiment, the vector is modified to allow transfer and integration of the polynucleotide into the host cell 5 genome.

Integration is a preferred way of ensuring stable expression. Integration can be obtained by physical recombination. Recombination can be homologous, e.g., between two homologous DNA sequences, or illegitimate, where recombination occurs between two 10 non-homologous DNAs. As a particular example, integration of donor DNA into the chromosome of the recipient can be mediated by host recombination repair system or by site-specific recombination. Another well-studied process that can transfer and integrate genes is transduction by bacterial viruses, such as λ and ϕ C31. In a phage-infected bacterial cell, fragments of the host DNA are occasionally packaged into phage particles 15 and can then be transferred to a recipient cell. Integration into the recipient cell's genome is caused by an integrase.

In a specific embodiment, the target polynucleotide construct comprises a nucleic acid 20 encoding an integrase functional in the selected host cell. More preferably, the integrase is selected from λ and ϕ C31 integrases. In a specific embodiment, the polynucleotide construct comprises a nucleic acid encoding an integrase having or comprising all or a 25 functional part of the following sequence of the ϕ C31 integrase (SEQ ID NO: 6) :

AGATCTCCGTACTGACGGACACACCGAAGCCCCGGCGAACCTCAGCGGATGC
25 CCCGGGGCTTCACGTTTCCCAGGTAGAAGCGGTTTCGGGAGTAGTGCCCCAACT
GGGGTAACCTTGAGTTCTCTCAGTTGGGGCGTAGGGTCGCCACATGACACAAG
GGGTTGTGACCGGGGTGGACACGTACGCGGGTGCTTACGACCGTCAGTCGCGAG
CGCGAGAATTCGAGCGCAGCAAGCCCAGCGACACAGCGTAGCGCCAACGAAGACA
AGGCGGCCGACCTTCAGCGCGAAGTCGAGCGCGACGGGGCGGTTCAAGGTTCGTC
30 GGGCATTTCAGCGAAGCGCCGGCACGTCGGCGTTCGGGACGGCGGAGCGCCCGGA
GTTCGAACCGCATCCTGAACGAATGCCGCGCCGGCGCTCAACATGATCATTGTCT
ATGACGTGTCGCCTCTCGCGCTGAAGGTCAATGGACCGCATCCGATTGTCTCGG
AATTGCTGCCCTGGCGTGACGATTGTTCACTCAGGAAGGCGTCTCCGGCAGG
GAAACGTCATGGACCTGATTCACCTGATTATGCGGCTCGACCGTCGACAAAGAA
35 TCTTCGCTGAAGTCGGCGAAGATTCTCGACACGAAGAACCTTCAGCGCGAATTGGG
CGGGTACGTCGGCGGGAAAGGCGCCTTACGGCTTCAGGCTTCAAGGAGACGAAGG

AGATCACGCGAACGGCGAATGGTCAATGTCGTCAACAAAGCTTGCACACTCG
5 ACCACTCCCCCTAACCGGACCCCTCGAGTCGAGCCGACGTAATCCGGTGGTGG
CGTGAGATCAAGACGCACAAACACCTTCAAGCGGGCAGTCAAGCCGCCAT
TCACCCGGGCAGCATCACGGGGCTTGTAAAGCGCATGGACGCTGACGCCGTGCCGA
10 CCCGGGGCAGACGATTGGGAGAAGACCGCTCAAGCGCTGGGACCCGGCAACC
GTTATGCGAATCTTCCGGACCCCGTATTGCGGGCTTCGCCGCTGAGGTGATCTAC
AAGAAGAAGCCGGACGGCACGCCGACCACGAAGATTGAGGGTACCGCATTAGCG
CGACCCGATCACGCTCCGGCCGGTCGAGCTGATTGCGGACCGATCATCGAGCCCG
15 CTGAGTGGTATGAGCTCAGGCCTGGTGGACGGCAGGGGGCGCGCAAGGGGCTT
TCCCGGGGCAAGCCATTCTGTCCGCATGGACAAGCTGTACTGCGAGTGTGGCGC
CGTCATGACTTCGAAGCGCGGGAAAGAATCGATCAAGGACTCTTACCGCTGCCGTC
GCCGGAAGGTGGTCGACCCGTCCGCACCTGGCAGCACGAAGGCACGTCAACGTC
AGCATGGCGGCACTCGACAAGTTCGTGCAGCATTCAACAAGATCAGGCA
20 CGCCGAAGGCAGAAGAGACGTTGGCGCTTGTGGGAAGCCGCCGACGCTTCG
GCAAGCTCACTGAGGCCTGAGAAGAGCGCGAACGGCGAACCTTGTGGAG
CGCGCCGACGCCCTGAACGCCCTGAAGAGCTGTACGAAGACCGCGGGCAGGC
GTACGACGGACCCGTTGGCAGGAAGCAGTCCGGAAGCAACAGGCAGCGCTGACGC
TCCGGCAGCAAGGGCGGAAGAGCAGCTTGCCTGAAGCCGCCAACGCCCCG
AAGCTTCCCCCTGACCAATGGTCCCCGAAGACGCCGACGCTGACCCGACCGCCCT
25 AAGTCGTGGTGGGGCGCGCTCAGTAGACGACAAGCGCGTTCGTCGGCTCTT
CGTAGACAAGATCGTTGTCAAGTCGACTACGGGAGGGGGCAGGGAACGCCA
TCGAGAAGCGCGCTCGATCACGTGGCGAACGCCGCCACCGACGACGAAGAC
GACGCCAGGACGGCACGGAAAGACGTAGCGCGTAGCGAGACACCCG

25 The term "functional part" designates any fragment or variants of the above sequence which retain the capacity to cause integration. Such fragments typically comprise at least 80%, preferably at least 85% or 90% of the above sequence. Variants may include one or several mutations, substitutions, deletions or additions of one or several bases.

30 In a more preferred variant, the target polynucleotide construct comprises genetic elements allowing transfer of the cloning vector into the selected host cell and integration of the cloning vector or a portion thereof into the genome of the selected host cell. Most preferred polynucleotide constructs comprises an oriT and a nucleic acid encoding an integrase.

35 In an other variant, the target polynucleotide construct comprises an origin of replication specific for or functional in the selected host cell. The origin of replication may be selected (or derived), for instance, from pAMβ1, pSa, 2μm circle, pSam2, pSG1, pIJ101, SCP2, pA387 and artificial chromosomes.

In an other variant, the target polynucleotide construct comprises a transcriptional promoter functional in the selected host cell. As indicated above, in a particular variant, the invention allows to modify the cloning vector to enable expression or over-expression of the cloned polynucleotides in the selected host. The expression of genes is 5 driven mainly by transcriptional promoters, which initiate gene transcription. The type of promoter to be used in the present invention can be selected by the skilled person, depending on the selected host cell and type of expression needed. Promoters may be ubiquitous or cell-specific, regulated or constitutive, weak or strong. They may be of various origins, including promoters isolated from viruses, phages, plant cells, bacterial 10 genes, mammalian genes, etc., or they may be artificial or chimeric. Typical examples of promoters include T7, T4, LacZ, trp, ara, SV40, tac, λ PL, GAL, AOX, hsp-70, etc.

The target polynucleotide construct is typically a DNA molecule, although RNAs may also be used as starting material. It is typically a double-stranded DNA. The target 15 polynucleotide construct may be produced by conventional recombinant DNA techniques, including DNA synthesis, cloning, ligation, restriction digestion, etc. and a combination thereof.

The target polynucleotide construct is preferably engineered so as to be inserted in a 20 region of the vector distinct from the polynucleotide. Indeed, it is important that the integrity of the polynucleotides is preserved. Directed insertion may be accomplished in a variety of ways, including site-specific insertion using particular enzymatic systems (Cre/Lox, FLP, etc.), homologous recombination with particular target sequences present in the vector, or by the use of transposons or transposable elements and appropriate 25 selection means.

In a particular, preferred embodiment, the target polynucleotide construct is contained in or comprises a transposable nucleic acid construct. Indeed, in a preferred variant, the methods of the present invention use transposable elements to alter the properties of the 30 cloning vectors, and allow their transfer, maintenance, expression or over-expression in a selected host cell.

Transposable nucleic acid constructs are derived from transposons, which are genetic elements capable of moving from one genetic loci to another. Two main classes of transposons have been identified in bacteria. The most simple transposons comprise an 5 insertion sequence that carries only elements of transposition. These elements are two inverted DNA repeats and a gene that codes for a protein called transposase. The transposase catalyses the excision and integration of the transposon. It has been shown that the excision and integration reaction can be catalysed in trans by a transposase, which can be provided in vivo or in vitro in purified form or expressed from a different 10 construct. More complex transposons carry more insertion sequences and additional genes that are not involved in transposition.

Transposable nucleic acid constructs of this invention thus typically comprise, flanked by two inverted repeats, the target polynucleotide construct and, more preferably, a 15 marker gene. In the presence of a transposase, these transposable nucleic acid constructs can integrate into a cloning vector in vivo or in vitro, thereby providing for targeted polynucleotide insertion. Alternatively, such nucleic acid constructs can be used for targeted integration, in the absence of a transposase, in particular strains such as hypermutator strains. Such transposable nucleic acid constructs also represent a 20 particular object of the present application. In this regard, in a more preferred embodiment, the invention also relates to a transposable nucleic acid construct, wherein said construct comprises an origin of transfer flanked by two inverted repeats. Specific examples of such construct are transposons pPL1 and pPAOI6, as disclosed in the experimental section. The transposable nucleic acid construct may further comprise an 25 integrase gene and/or a marker gene.

The inverted repeat nucleic acid sequences may be derived from the sequence of various transposons, or artificially created. In particular, transposable elements can be generated using inverted repeats obtained from transposons or transposable elements such as Tn5, 30 Tn21, miniTn5, T7, T10, Tn917, miniTn400, etc. Preferably, the sequences derive from

transposon Tn5. In a specific embodiment, they comprise all or a functional part of the following sequences :

- left arm of pPAOI6 transposon (SEQ ID NO: 7)

5 CTGTCTCTTATACACATCTAACCATCATCGATGAATTTCGGGTGTTCTCGCATA
TTGGCTCGAATTGAGCTCGGTACCC

- right arm of transposon pPAOI6 (SEQ ID NO: 8)

10 GATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGCCAACGACTACGCACTAGCC
AACAAAGAGCTTCAGGGTTGAGATGTGTATAAGAGACAG

The marker gene may be any nucleic acid encoding a molecule whose presence in a cell can be detected or visualized. Typical marker genes encode proteins conferring resistance to antibiotics, such as apramycine, chloramphenicol, ampiciline, kanamycine, 15 spectinomycine, thiostrepton, etc. Other types of markers confer auxotrophy or produce a label (e.g., galactosidase, GFP, luciferase, etc).

In a specific embodiment, the cloning vector in the library comprises a first marker gene and the modification step ii) comprises:

20 . contacting in vitro, in the presence of a transposase, the selected cloning vectors with a transposon comprising, flanked by two inverted repeats, the target polynucleotide construct and a second marker gene distinct from the first marker gene, and
. selecting the cloning vectors which have acquired the second marker gene and which have lost the first marker gene.

25

The double selection ensures that the target polynucleotide construct has been inserted at a site within the marker gene present in the cloning vector, i.e., outside of the polynucleotide insert.

30 It should be understood that the modification may be accomplished in various other ways, particularly by incorporating a sequence coding for the transposase directly into the transposable element or into another expression unit. The presence and expression of the transposase can be regulated by inductive promoter or termosensitive replicative units. Also, transposition can be carried out by *in vitro* process.

Analysis of the polynucleotides

In step (iii) of the process, the polynucleotides may be analysed by various methods, including by genetic, biochemical, chemical or phenotypical approaches, which are well known per se in the art. Analysis occurs upon transfer and, optionally, expression of the polynucleotides into the selected host cell.

In this regard, the modified cloning vectors can be transferred into the selected host cell by a variety of techniques known in the art, including by transformation, electroporation, transfection, protoplast fusion, conjugative transfer, etc. In a preferred embodiment, the target polynucleotide construct comprises an oriT and the modified vectors are transferred into the selected host cells by conjugative transfer. In this embodiment, the cloning vector and the selected host cells are co-cultivated and the recombinant host cells are selected and isolated.

The selected host cell may be any type of cell or microorganism, including, without limitation, *Streptomyces*, *E. coli*, *Salmonella*, *Bacillus*, *Yeast*, *fungi*, etc.

One of the objectives of the invention is to be able to analyse environmental DNAs of unknown cellular origin into different host expression systems. In order to analyse the potentiality of the DNAs at the transcription and/or translation levels and to have much more probabilities to have a DNA expression, it is important to have the possibility to test different host expression systems.

Insertion of foreign DNA into host expression systems like *Streptomyces* can produce, for instance, an increase in doubling time, morphological modifications, pigments production, etc., which can be related either directly to the expression of foreign DNA or by combinatorial biology of the foreign DNA and the biology of the expression host systems. The new phenotypes can be analysed by all techniques known in the art such genetic, biochemical, chemical, phenotypic approaches etc.

In a preferred, specific embodiment, the invention relates to a method for the identification or cloning of polynucleotides encoding a selected phenotype, the method comprising (i) cloning environmental DNA fragments into *E. coli* cloning vectors to 5 produce a metagenomic library, (ii) identifying or selecting cloning vectors in said library which contain DNA fragments having a particular characteristic of interest, (iii) modifying the identified or selected cloning vectors into shuttle or expression vectors for transfer and integration in a selected host cell, (iv) transferring the modified cloning vectors into said selected host cell and (v) identifying or cloning the DNA fragments 10 contained in said modified cloning vectors which encode said selected phenotype in said selected host cell.

By applying the above method, new polynucleotide sequences have been identified, cloned and characterized, which produce new phenotypes in bacteria. These 15 polynucleotides contain the sequence of PKS genes and other genes that encode polypeptides involved in biosynthetic pathways. The sequence of these polynucleotides is provided in Figure 5. The invention also relates to any polynucleotide sequence comprising all or part of these sequences (i.e., SEQ ID NOS: 1 or 2), their complementary strand, or a functional variant thereof. A part of the above sequences 20 includes, preferably, at least 20 consecutive bases, more preferably at least 50 consecutive bases thereof, even more preferably a coding sequence (e.g., an CDS). In this respect, SEQ ID NOS: 1 and 2 comprise several novel open reading frames encoding novel polypeptides involved in biosynthetic pathways. These coding sequences are identified in Figures 5a and 5b.

25

In a specific embodiment, the invention relates to a polynucleotide sequence comprising a sequence selected from nucleotides (CDS) 76 - 1134 ; 1096 - 2430 ; 1178 - 1624 ; 2506 - 3567 ; 2906 - 4222 ; 4092 - 5321 ; 6337 - 8502 ; 8181 - 9530 ; 9531 - 10721 ; 10504 - 11274 ; 12874 - 13689 ; 14195 - 15976 ; 15427 - 16512 ; 15579 - 16253 ; 16505 30 - 17656 ; 17657 - 18697 ; 18615 - 19304 ; 19301 - 20596 ; 20535 - 21476 ; 22025 -

22951 ; 23155 - 26523 ; 26409 - 34433 ; 34418 - 37500 and 35359 - 37500 of SEQ ID NO: 1 (figure 5a) or a complementary strand thereof.

In an other specific embodiment, the invention relates to a polynucleotide sequence 5 comprising a sequence selected from nucleotides 3 - 914, 924 - 2168 ; 2207 - 3190 ; 3373 - 4455 ; 4546 - 4959 ; 5176 - 6192 ; 6331 - 14043 ; 14275 - 15408 ; 15436 - 16245 ; 16287 - 17384 ; 17427 - 18158 ; 18248 - 18847 ; 18952 - 20346 ; 20442 - 21167 ; 21164 - 24301 ; 24351 - 27023 ; 27806 - 29686 ; 29535 - 30872 ; 30848 - 32647 ; 32574 - 35555 ; 35533 - 36598 and 36516 - 37400 of SEQ ID NO: 2 (figure 5b) or a 10 complementary strand thereof.

Variants of these sequences include any naturally-occurring variant comprising or or 15 several nucleotide substitutions ; sequences variants resulting from the degeneracy of the genetic code, as well as synthetic variants coding for functional polypeptides. Variants include any sequence that hybridise under high stringent conditions, as disclosed for instance in Sambrook et al., to any of the above sequences, and encode a functional 20 polypeptide. The invention also include any nucleic acid molecule encoding a polypeptide comprising all or a fragment of an amino acid sequence encoded by a polynucleotide as disclosed above. Preferably, the fragment comprises at least 10 consecutive amino acid residues, more preferably at least 20, even more preferably at least 30.

The invention also relates to any vector comprising these polynucleotide sequences. 25 These sequences may be DNA or RNA, preferably DNA, even more preferably double-stranded DNA. The invention also relates to a polypeptide encoded by a polynucleotide sequence as defined above. The invention also relates to a method of producing such polypeptides by recombinant techniques, comprising expressing a polynucleotide as defined above in any suitable host cell and recovering the encoded polypeptide. The invention also relates to a recombinant host cell comprising a polynucleotide or a vector 30 as defined above.

An other object of this invention resides in a library of polynucleotides, wherein said library comprises a plurality of environmental DNA fragments cloned into cloning vectors, wherein said environmental DNA fragments contain a common molecular characteristic and wherein said cloning vectors are *E. coli* cloning vectors comprising a target polynucleotide construct allowing transfer and integration of the environmental DNA into the genome of a selected host cell distinct from *E. coli*.

The sub-DNA-libraries should have either desired genetic characteristics based on high or low GC content, DNA encoded for a desired enzymatic activity, part or full biosynthetic pathways for metabolites etc., or specific origin such as soil fractions, 10 animal organs, sub fraction of a microorganism community etc. The invention also allows to produce conjugative vector with desired characteristics in accordance with the characteristics of the pre-identified sub-DNA-libraries and functional analysis of mutants in heterologous hosts. It can also be used, without limitation, for the production of 15 mutants by mutagenesis, for DNA sequencing, genes or biosynthetic pathways knock-out by insertion or to confer transfer capabilities for expression, co-expression, over-expression or modification of biosynthetic pathways.

Further aspects and advantages of the invention will be disclosed in the following 20 examples, which should be regarded as illustrative and not limiting the scope of this application.

Experimental Section

25 A - From *E.coli* to *Streptomyces*

In this work, we constructed a fosmid library in *E. coli* from total DNA prepared directly from soil. The library has been screened for presence of biosynthetic pathways. We developed genetic tools for functional genomics that allow gene identification, 30 inactivation and horizontal gene transfer from *E. coli* to *Streptomyces*.

The cloning vectors in the library contain ColE1 replicon for propagation in *E. coli*. Transposable elements based on Tn 5 transposon were produced and used for *in vitro* modification of selected cloning vectors. Integrated transposable elements contain gene for resistance to apramycine. Conjugative derivatives were constructed by incorporating 5 origin of transfer from plasmid RP4. A conjugative and site specific integrative transposon was also constructed comprising the integrase gene from ϕ C31 phage, including attP attachment site. Conjugal transfer was demonstrated from an appropriate *E. coli* donor cell to another *E. coli* or *Streptomyces lividans* recipient cell. Constructed transposon was tested for inactivation of the genes cloned into fosmids. Obtained 10 mutants can be used for direct sequencing by adequate primers and transferred by conjugation into *Streptomyces lividans*. Transposable elements thus represent very useful tools for functional analysis of a large DNA libraries cloned into BAC, PAC fosmids or other cloning vectors in which cloned inserts must be transferred into heterologous host.

Materials and methods

15 Bacterial strains, plasmids and growth conditions.

E. coli DH10B (F- mcrA delta(mrr-hsdRMS-mcrBC) phi80dlacZ deltaM15 delta lacX74 deoR recA1 endA1 araD139 delta (ara, leu)7697 galU galK lambda- rpsL nupG), strain (Epicentre) was used for fosmid and plasmid transformation and DNA amplification. Unless specifically described, all DNA manipulations were performed according to 20 Sambrook, J., *et al*, (1989).

Soil DNA extraction and DNA libraries construction.

Total bacterial community DNA was extracted and large DNA libraries have been constructed into fosmids according to the method described in WO 01/81357.

Fosmid DNA extraction and Purification

25 Fosmids DNA containing soil library were extracted from pools of 96 clones. Culture of recombinant clones were performed in Deep-Well 96 and 48, respectively in 1 ml and 2 ml of LB media containing 12,5 μ g ml⁻¹ of chloramphenicol. Cultures were grown at

37°C with shaking at 250 RPM during 22 hours. DNA extraction was done by using the Nucleobond PC100 extraction kit (Macherey Nagel).

PCR screening for the detection of PKS genes

5 Primers design

Degenerate PCR primers sets were designed to specifically amplify PKS nucleic acids sequences. Multiple sequence alignment of PKS domains revealed highly conserved motives, in particular in the β -keto acyl synthase domain. Primers Lib1F and Lib2R were designed in conserved motives of the β -keto acyl synthase gene. Lib1F (sense primer, 5'- GGSCCSKCSSTSDCSRTSGAYACSGC -3') and Lib2R (antisense primer, 5'- GCBBSRYYTCDATGGRTCSGC -3') were deduced from β -keto acyl synthase peptide sequences GP(AS)(LV)(AST)(IV)DTAC and GDPIE(TVA)(RAQ)A, respectively. The specific fragment amplified with Lib1F / Lib2R was approximately about 465 bp (corresponding to 155 amino acids). Specificity and efficiency of the PCR systems were validated by testing on positive DNA controls (i.e. genomic DNA from type I PKS producing strain such as *Bacillus subtilis*, *Streptomyces lividans*, *Streptomyces ambofaciens* and *Ralstonia solanacearum*) and negative DNA controls (genomic DNA from strains which are known to do not contain PKS genes). Furthermore, DNA extracted from soil samples were tested to calibrate PCR techniques.

20

PCR conditions

PCR conditions were optimised, in particular for concentrations of DMSO, MgCl₂, Primers and DNA template quantities. For PCR using microorganism genomic DNA and soil DNA as template (50 to 200 ng), the PCR mix (50 μ l) contained 250 μ M of dNTP, 5 mM MgCl₂ final, 2,5% DMSO, 1X PCR buffer, 0,75 μ M of each primer and 2,5 U of Taq DNA polymerase (Sigma) and sterile distilled water. For PCR using fosmid pooled DNA as template (100 to 500 ng), the PCR mix (50 μ l) contained 250 μ M of dNTP, 5 mM MgCl₂ final, 5% DMSO, 1X PCR buffer, 0,75 μ M of each primer and 2,5 U of Taq DNA polymerase (Sigma) and sterile distilled water. For identification of positives clones in 96 microtiter plates, 25 μ l of each bacteria culture were used as template and PCR conditions were the same as above. Thermocycling program was : a denaturation

step at 96°C for 5 minutes; then 1 minute at 96°C, 65°C for 1 minute, 72°C for 1 minute. The first 7 cycles, the annealing temperature was lowered 1°C per cycle until 58°C was reached. A subsequent 40 cycles were carried out with the annealing temperature at 58°C. A final extension step was at 72°C for 7 minutes. For identification of positives 5 clones in 96 microtiter plates, the first denaturation step of 96°C was during 8 minutes. The other steps were the same as described above. PCR reactions were performed with a PTC 200 thermocycler (MJ Research).

PCR products analysis

10 PCR products of about 465 bp were purified on agarose gel with gel extraction Kit (Qiagen) according to the manufacturer recommendations. First approach consisted in subcloning PCR products using the Topo PCR II kit (Invitrogen). Recombinant Plasmids were extracted using QIAprep plasmid extraction Kit (Qiagen) and sequenced with Forward and Reverse M13 primers with CEQ 2000 automated sequencer (Beckman 15 Coulter). Second approach consisted in direct sequencing of PCR products. Sequencing data were compared with nucleic and proteic genbank database using BLAST program.

Sequencing of the identified fosmid insert DNA and sequence analysis

20 Fosmids inserts were sequenced using either a transposon-mediated and by shotgun subcloning approach. Transposition was realized by using (Transposition Kit) commercialized by Epicentre according to the manufacturer. For shotgun subcloning, transformants were grown for 16 hours at 37°C. Fosmid extraction was done by using the Nucleobond PC100 extraction kit (Macherey Nagel). DNA was partially restricted with *Sau3A* and sized on standard gel electrophoresis for fragments ranging from 1 to 3 25 Kbs and cloned into Bluescript vector according to Sambrook *et al.* (1989).

Sequence analysis was performed with the identification of ORFs by using Frameplot of the GC3. Each identified ORF was compared to gene databases by using BLAST program. PKS domains were determined by aligning obtained sequence versus already described PKS domains from domain databases.

Sequencing

Sequencing reactions were performed with 1 µg of DNA and 3.6 pmol of primer, using CEQ 2000 Dye Terminator Cycle Sequencing kit (Beckman Coulter) under conditions proposed by supplier. Ten µL of reaction products were precipitated using 4 µL of solution containing 1.5 M NaOAc, 50 mM EDTA and 60 µL cold 95% ethanol/dH₂O from -20 °C. The pellet was washed 2 times with 200 µL 70% ethanol/dH₂O, vacuum dried and dissolved in 40 µL sample loading solution (supplied in kit). Sequencing reactions were run on an CEQ 2000 sequencer (Beckman Coulter).

10 Plasmids construction and validation

Plasmid pP1 was constructed as follows. A 941 bp DNA fragment containing native promoter region and AA(3)IV gene was amplified by polymerase chain reaction (PCR) using primers AmF (d-CCCTAAGATCTGGTCATGTGCAGCTCCATC, SEQ ID NO: 9) and AmR (d-TAGTACCCGGGGATCCAACGTCATCTCGTTCTCC, SEQ ID NO: 10). One hundred microlitre reaction were performed containing 0.1µM each of primers, 1 X Vent DNA polymerase buffer (NEB), 0.2 µM of each deoxyribonucleoside triphosphate (dNTP), 50 ng of the DNA template and 2U of Vent DNA polymerase (NEB). PCR mixture was heated for 4 min at 94 °C in a PTC-200 thermocycler (Peltier) and cycled 25 x at 94 °C for 60 sec, 59 °C for 30 sec and at 72 °C for 70 sec. The final 20 extension was performed at 72 °C for 7 min.

PCR product was purified using GFX DNA purification kit (Amersham), then digested by *Bgl* II and *Sma* I restriction enzymes. A 941 bp *Bgl*II/*Sma*I fragment was inserted into the *Bam* HI, *Sma* I sites of pMOD plasmid (EPICENTRE). DH10B *E. coli* was transformed with pP1 and subjected to apramicyne selection on LB agar plates. Six 25 colonies surviving on apramicyne selection were grown in liquid LB media and final pPL1 candidates were thoroughly checked via PCR and restriction mapping.

To construct conjugative plasmid pPL1, 750 bp oriT DNA region from plasmid RP4 were amplified via PCR using primers oriTF (d-
30 GCGGTAGATCTGTGATGTACTTCACCAGCTCC, SEQ ID NO: 11) and oriTR

(TAGTACCCGGGGATCCGACGGATCTTTCCGCTGCAT, SEQ ID NO: 12). PCR conditions were as above. Amplified DNA was digested using *Bgl* II and *Sma*I restriction enzymes. *Bgl* II/*Sma*I DNA fragment was subjected to purification after gel electrophoresis on 0.7% agarose. Purified fragment was ligated into pP1 plasmid digested by *Bam* HI and *Sma* I restriction enzymes.

5 ϕ C31 integrase gene and attachment site (attP) was amplified via PCR using primers Fint (d-AACAAAGATCTCCGTACTGACGGACACACCG, SEQ ID NO: 13) and PJ (d-CGGGTGTCTCGCATGCCGCT, SEQ ID NO: 14). Amplified DNA fragment was purified by GFX kit (Amersham) and phosphorilated using T4 polynucleotide kinase (NEB) under conditions recommended by the enzyme manufacturer. Phosphorilated DNA fragment was cloned in to pPL1 vector opened with *Sma*I restriction enzyme (NEB) and dephosphorilated by calf alkaline phosphatase (NEB). DH10B *E. coli* was transformed with ligation mixture using Bio Rad Pulsing apparatus and protocols 10 provided by Bio-Rad. Twelve transformants were analyzed by PCR for the presence of integrase gene. Orientation of integrase gene was verified by restriction analysis using *Bgl* II and *Eco*RI restriction enzymes. Resulting plasmid was named pPAOI6.

15 To construct pPAOI6-A plasmid, pPAOI6 plasmid was digested with *Eco*RI and *Bgl* II restriction enzymes followed by digestion by Bean mung nuclease (NEB). Linearised 20 plasmid was self ligated and transformed in to DH10B cells.

Plasmid preparation of fosmid DNA

Fosmid and BAC DNA for sequencing was prepared by using the Nucleobond AX kit (Macherey-Nagel), following protocol for BACs, Cosmid as specified by manufacturer.

Mutagenesis

25 Transposon Tn-pPAOI6 was prepared by digestion of pPAOI6 plasmid using *Pvu*II restriction enzymes, followed separation on agarose gel and purification of fragment containing transposon from gel using Qiagen kit. The same molar ratio of transposon and corresponding fosmid was used for mutagenesis in vitro using Tn5 transposase (Epicentre) and conditions specified by manufacturer. We transformed aliquot of the 30 transposed mixture by electroporation into competent DH10B *E. coli* strain.

Conjugation *E. coli*- *Streptomyces lividans* TK24

Conjugation experiments were done using 6×10^6 *E. coli* S17.1 cells containing conjugative plasmids or fosmids. The *E. coli* cells were grown in LB media with adequate antibiotic. The cells were collected by centrifugation, washed two times using same volume of LB media and concentrated to 10^8 cells/ml and overlaid on LB plates containing 2×10^6 pregerminated *Streptomyces lividans* TK24 spores. The cells mixture were grown over night at 30^0 C and *E. coli* cells were washed three times using 2 ml of LB media. The plates were overlayed using top agar containing NAL (nalidixic acid) and the appropriate antibiotic. Plates were incubated for 4 days at 30^0 C and transformant streptomyces colonies were isolated on HT medium (Pridham *et al.* 1957) containing NAL and the same appropriate antibiotic.

Results**Construction of a Transposon Tn <Apra>.**

15 *E. coli* aminoglycoside -(3)- acetyl transferase IV gene (aa(3)IV) was amplified by PCR and cloned in pMOD vector (Epicentre) . Advantage of this selective marker allows positive selection in *E. coli* and in *Streptomyces lividans*. Transposon can be used for insertional inactivation in vitro using purified transposase Tn5 (Epicentre). The structure of the pP1 constructed vector and transposon was shown on Figure 1a,1b.

20 Construction of a conjugative plasmid-transposon.

Conjugative vector, transposon was constructed by cloning origin of transfer from plasmid RP4 into pP1 vector producing pPL1 vector (Figure 2a, 2b). The origin of transfer was cloned in such orientation that the selective aa(3)IV gene is the last transferred during conjugation. PPL1 vector was introduced into specific *E. coli* S17.1 25 strain that carry RP4 plasmid integrated into chromosome. In conjugation experiment between donor strain S17.1 carrying pPL1 plasmid and DH10 *E. coli* receptor we obtained DH10B strain carrying pPL1 plasmid. This data shows that cloned oriT fragment is functional in the pPL1 plasmid. Plasmid pPL1 can be used for DNA cloning, gene inactivation by homologous recombination. Cloned genes or part of the gene cloned

could be then transferred by conjugation in to another host. Another advantage of this vector is conjugative transposon that can be excised from vector and inserted randomly in vivo in to another DNA molecule by purified Tn 5 transposase.

Construction of a conjugative, site specific integrative plasmid- transposon for horizontal
5 gene transfer between *E. coli* and *Streptomyces* strains.

PPL1 plasmid was used to clone an integrase gene from phage ϕ C31, resulting a plasmid pPAOI6 (Figure 3a, 3b). We tested several clones for horizontal transfer between *E. coli* S17.1 strain and *Streptomyces lividans* TK24 strain. The best transfer was obtained for plasmid pPAOI6 were orientation of integrase gene is in opposite orientation to the gene
10 for resistance to apramycine. Conjugative transfer of pPAOI6 gene in to *S. lividans* strain is confirmed by resistance to apramycine or G418. Additional confirmation of transfer was obtained using PCR method. We were able to amplify 2 kb insert using specific primers for the ϕ C31 integrase gene, and no PCR amplification was obtained for control *S. lividans* TK24 strain.

15 pPAOI6 transposon was cloned into the *EcoRV* site of plasmid pGPS3 (New England Biolabs). This construction allows transposition not only by transposase Tn5 but also using Transposase ABC (New England Biolabs). Resulting plasmid pTn5-7AOI is shown on Figure 10.

The goal of these constructions was to produce transposons that is further used in
20 functional analysis of the metagenomic DNA library from the soil that were constructed in a laboratory (Figures 4a, 4b, 4c).

Functional analysis of the metagenomic DNA library from soil.

The fosmids library consists of 120 512 clones, containing ~ 40 kb inserts of soil DNA.

The library contains approximately 4.8 Gbps of the DNA cloned from soil. Ten percents
25 of the library was screened by using a PCR approach for the presence of the genes

involved in production of secondary metabolites (PKS). Using gene-module specific set of primers we were able to identified positive clones organized in microtiter plates (96 wells). Sequences (based on PCR products) obtained from fifteen randomly positive clones indicate that the DNA library contains very little sequence redundancy limited to

one and that the sequences were found to be new and very diverse in comparison to gene databases (data not shown).

Two fosmids DNA was prepared from two positive clones (FS3-124 and FS3-135) and analyzed by sequencing. DNA analysis *in silico* shows high G+C contents of 72% and 5 69% respectively of the cloned inserts and presence of cluster genes that could be involved in biosynthesis of secondary metabolites (Fig 5 a and b). No specific phenotype was observed for the two clones in *E. coli*. We employed pPAOI6 transposon mutagenesis to produce conjugative mutants. Transposon mutants FS3-124::pPAOI6 and FS3-135::pPAOI6 were isolated using apramycin as selective antibiotic. Obtained 10 mutants were then tested on LB plates containing chloramphenicol. About 1% of the tested clones are chloramphenicol sensitive. These clones contain transposon inserted into locus encoding chloramphenicol resistance gene and not into cloned DNA insert. ApraR and ChloS transposon mutants are then used for horizontal gene transfer into *Streptomyces lividans* TK24 strain. Fosmid DNA was prepared from mutants and 15 transformed into *E. coli* S17.1 strain. Horizontal gene transfer between *E. coli* S17.1 and *Streptomyces lividans* was done due to inserted pPAOI6 transposon. Transconjugants of the *Streptomyces lividans* were tested by PCR to confirm gene transfer and integration of the conjugative fosmid in to *S. lividans* chromosome. Both transconjugants showed an increase in doubling time, morphological modifications and pigments production in 20 comparison to the control (Figure 6).

B - From *E. coli* to *Bacillus subtilis*

Plasmid pPSB was constructed as follows : A part of amyE gene from *B. subtilis* was 25 amplified by PCR using primers amyE-BamHI : atcgcaggatccctgaggactctcgaaacccg (SEQ ID NO: 15) and amyE-EcoRI : cgactgaattcagatctagcgtgtaaattccgtctgc (SEQ ID NO: 16). DNA fragment was digested by EcoRI and BamHI restriction enzymes and ligated into EcoRI, BglII site of pPAOI6 plasmid. Transposable element contains Tn<amyE-int φC31-oriT-apra> is shown on figure 7 with plasmid pPSB.

Plasmid pPSBery (Figure 8) was constructed by cloning erm AM gene into pPSB plasmid. A 1140 bp *Sau*3A DNA fragment containing ery AM gene with his own promoter was cloned from plasmid pMUTIN (Wagner *et al.* 1998) into the Bam HI site of plasmid pPSB. Orientation of ery gene was confirmed by sequencing. Transposable 5 element contains Tn < amyE-int ϕ C31-ery-oriT-apra> (Figure 8).

Plasmid pPSBery- Δ I was obtained after *Sma* I digestion and self-ligation of core plasmid (Figure 9). In this construction, ϕ C31 integrase gene was deleted from transposable element. New transposon is Tn<amyE-ery-oriT-apra> (Figure 9). All transposons could 10 be released as linear DNA by *Pvu*II digestion from plasmids mentioned above and transposed by transposase Tn 5 (Epicentre) in vitro.

The selection of transposed elements was done using 100 μ g/ml erythromycine or 40 μ g/ml apramycine in *E. coli* or 0.3 μ g/ml erythromycine in *B. subtilis*. DNA was 15 transformed by electro-transformation into electrocompetent *E. coli* strains or by competence into *B. subtilis*. Integration of imported DNA into amy E locus of *B. subtilis* chromosome was confirmed using pPSBery- Δ I plasmid. Integration was confirmed by PCR using plasmid-specific and amyE locus-specific primers. Fifteen eryR *B. subtilis* clones were tested by PCR. All transformants showed integration at amyE locus of *B. subtilis* chromosome, confirming the functionality of the method and constructs. 20

References

Brau, B., Pliz, U., and Pipersberg, W. 1984. Genes for gentamicin-(3)-N-acetyltransferases III and IV: I. Nucleotide sequence of the AAC(3)-IV gene and possible 5 involvement of an IS 140 element in its expression. *Mol Gen Genet.* 193 (1), 179-187.

Guiney, D.G., Yakobson, E., (1983). Location and nucleotide sequence of the transfer origin of the broad host range plasmid RK2. *Proc Natl Acad Sci USA*, 80: 3595-8.

10 T. Kaeberlein, K. Lewis, Epstein S. S. 2002. Isolating "Uncultivable" Microorganisms in Pure Culture in a Simulated Natural Environment. *Science* 296, 1127-1129.

Mazodier, P., Thompson, C., 1989. Intergeneric conjugation between *E. coli* and *Streptomyces* species. *J. Bacteriol.* 171, 3583-3585.

15 Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. *Science* 276: 734-740.

Pridham, T.G., Anderson P., Foley, C., Lindenfelser, L.A., Hesseltine, C.W., and Benedict R.C. (1957). A selection of media for maintenance and taxonomic study of 20 *streptomyces*. *Antibiotics Annual 1956-1957*, 947-953.

Voeykova T, Emelyanova L, Tabakov V, Mkrtumyan N. 1998. Transfer of plasmid pTO1 from *Escherichia coli* to various representatives of the order *Actinomycetales* by intergeneric conjugation. *FEMS Microbiol Lett.* 162(1):47-52.

25 Sambrook, J., Fritsch E.F. and Maniatis T. 1989. Molecular cloning: a laboratory manual, 2nd edition. Cold Spring Harbor Laboratory, Cold Spring Harbor NY.

Torsvik, V., Goksoyr, J. Daae, F. L. 1990. High diversity in DNA of soil bacteria. *Appl. Environ Microbiol.* 56, 782-787.

Torsvik, V., Sorheim, R. and Goksoyr, J. Total bacterial diversity in soil and sediment communities a review. *J. Industr. Microbiol.* 17, 170-178.

5 Torsvik, V., Daae, F.L, Sandaa, R.A., Overas, L. 1998. Novel techniques for analysing microbial diversity in natural and perturbed environments. *J. Biotechnol.* 64, 53-62.

Wagner, V., Dervyn, E. and Ehrlich S. D. 1998. A vector for systematic gene inactivationin *Bacillus subtilis*. *Microbiology*. 144, 3097 – 3104.

10 Zaehner, H. and Fiedler, F. P. 1995. Fifty years of antimicrobials: past perspectives and future trends. In The need for New antibiotics: Possible Ways Forward. Fifty –Third Symposium of the Society for General Microbiology, Bath. England, UK, (Hunter, P.A., Darby, G.K. & Russell, N.J., eds), pp. 67-84.

15 Zechner, E.L., de la Cruz, F., Eisenbrandt, R., Grahn, A.M., Koraimann, G., Lanka, E., Muth, G., Pansegrouw, W., Thomas, C.M., Wilkins, B.M. and Zatyka, M. (2000) Conjugative-DNA transfer process. In Thomas, C.M. (ed.) The horizontal gene pool. Harwood Academic Publishers.

CLAIMS

1. A method of analysing a library of polynucleotides, said polynucleotides being contained in cloning vectors having a particular host range, the method comprising (i) 5 selecting cloning vectors in the library which contain a polynucleotide having a particular characteristic, (ii) modifying said selected cloning vectors to allow a transfer and integration of said vectors and/or of the polynucleotide which they contain into a selected host cell, and (iii) analysing the polynucleotides contained in said modified vectors upon transfer of said modified vectors into said selected host cell.
- 10 2. The method of claim 1, wherein the library comprises a plurality of unknown polynucleotides.
- 15 3. The method of claim 1 or 2, wherein the library comprises a plurality of environmental DNA fragments.
4. The method of any one of the preceding claims, wherein the cloning vectors of the library are *E. coli* cloning vectors, preferably cosmids, fosmids, P1 or BAC.
- 20 5. The method of any one of the preceding claims, wherein the selected vectors are modified by targeted insertion, into the vector, of a target polynucleotide construct.
6. The method of claim 5, wherein the targeted insertion is performed in a region of the vector distinct from the polynucleotide.
- 25 7. The method of claim 5 or 6, wherein the target polynucleotide construct comprises an origin of transfer functional in the selected host cell.
8. The method of claim 7, wherein the origin of transfer is functional in *E. coli* host cells.

9. The method of claim 8, wherein the origin of transfer is selected from RP4, pTiC58, F, RSF1010, ColE1 and R6K(α).
10. The method of any one of claims 5 to 9, wherein the target polynucleotide construct
5 comprises an integrase functional in the selected host cell.
11. The method of claim 10, wherein the integrase is selected from φC31.
12. The method of any one of claims 6 to 11, wherein the target polynucleotide construct
10 comprises a transcriptional promoter functional in the selected host cell.
13. The method of any one of claims 5 to 12, wherein the target polynucleotide construct
comprises a transposable nucleic acid construct.
14. The method of claim 13, wherein the transposable nucleic acid comprises, flanked by
two inverted repeats, the target polynucleotide construct and a marker gene.
15. The method of any one of the preceding claims, wherein the cloning vector
comprises a first marker gene and wherein, in step ii), the cloning vector is modified by:
20 . contacting in vitro, in the presence of a transposase, the selected cloning vectors with a
transposon comprising, flanked by two inverted repeats, the target polynucleotide
construct and a second marker gene distinct from the first marker gene, and
. selecting the cloning vectors which have acquired the second marker gene and which
have lost the first marker gene.
16. The method of any one of the preceding claims, wherein, in step (i), the cloning
vectors which contain a polynucleotide having a particular characteristic are selected by
molecular screening.
17. The method of any one of the preceding claims, wherein, in step (iii), the modified
cloning vectors are transferred into the selected host cell by conjugative transfer.

18. The method of any one of the preceding claims, wherein, in step (iii), polynucleotides are analysed by determining the phenotype or properties of the host cell upon transfer or expression of the modified vector.

5

19. A method for the identification or cloning of polynucleotides encoding a selected phenotype, the method comprising (i) cloning environmental DNA fragments into *E.coli* cloning vectors to produce a metagenomic library, (ii) identifying or selecting cloning vectors in said library which contain DNA fragments having a particular characteristic of interest, (iii) modifying the identified or selected cloning vectors into shuttle or expression vectors for transfer and integration in a selected host cell, (iv) transferring the modified cloning vectors into said selected host cell and (v) identifying or cloning the DNA fragments contained in said modified cloning vectors which encode said selected phenotype in said selected host cell.

10 15

20. A transposable nucleic acid construct, wherein said construct comprises an origin of transfer and elements for integration and selection in a selected host cell genome flanked by two inverted repeats.

20 21. A library of polynucleotides, wherein said library comprises a plurality of environmental DNA fragments cloned into cloning vectors, wherein said environmental DNA fragments contain a common molecular characteristic and wherein said cloning vectors are *E. coli* cloning vectors comprising a target polynucleotide construct allowing transfer and integration of the environmental DNA into a selected host cell distinct from *E. coli*.

25 22. A polynucleotide sequence comprising all or part of SEQ ID NOs: 1 or 2, or of their complementary strand.

30 23. An oligonucleotide comprising SEQ ID NO: 3 or 4.

1/39

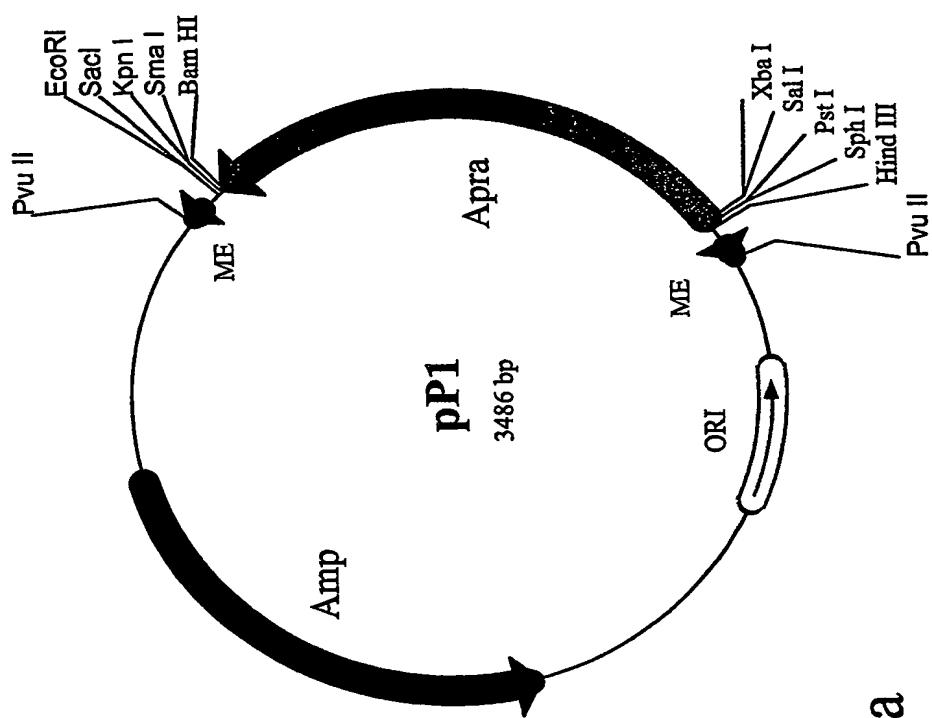


Figure 1a

2/39

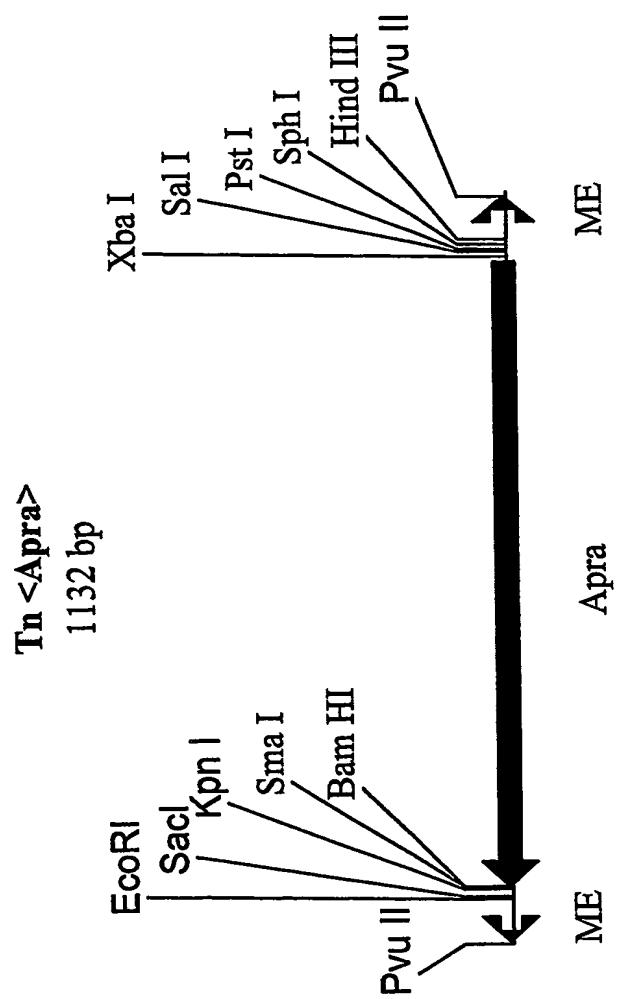


Figure 1b

3/39

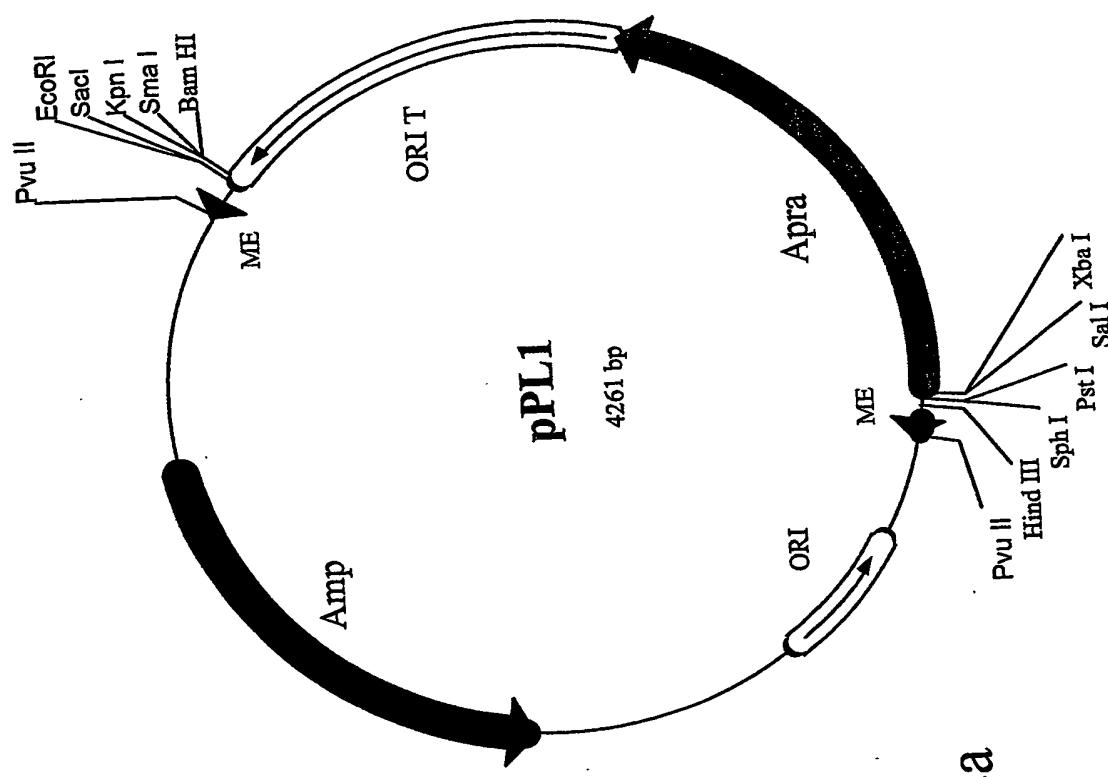


Figure 2a

4/39

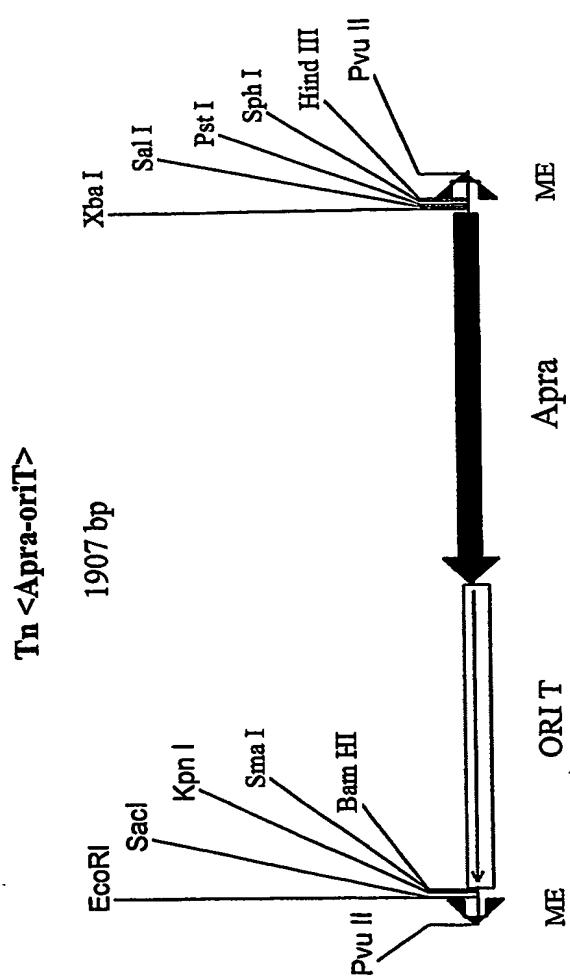


Figure 2b

5/39

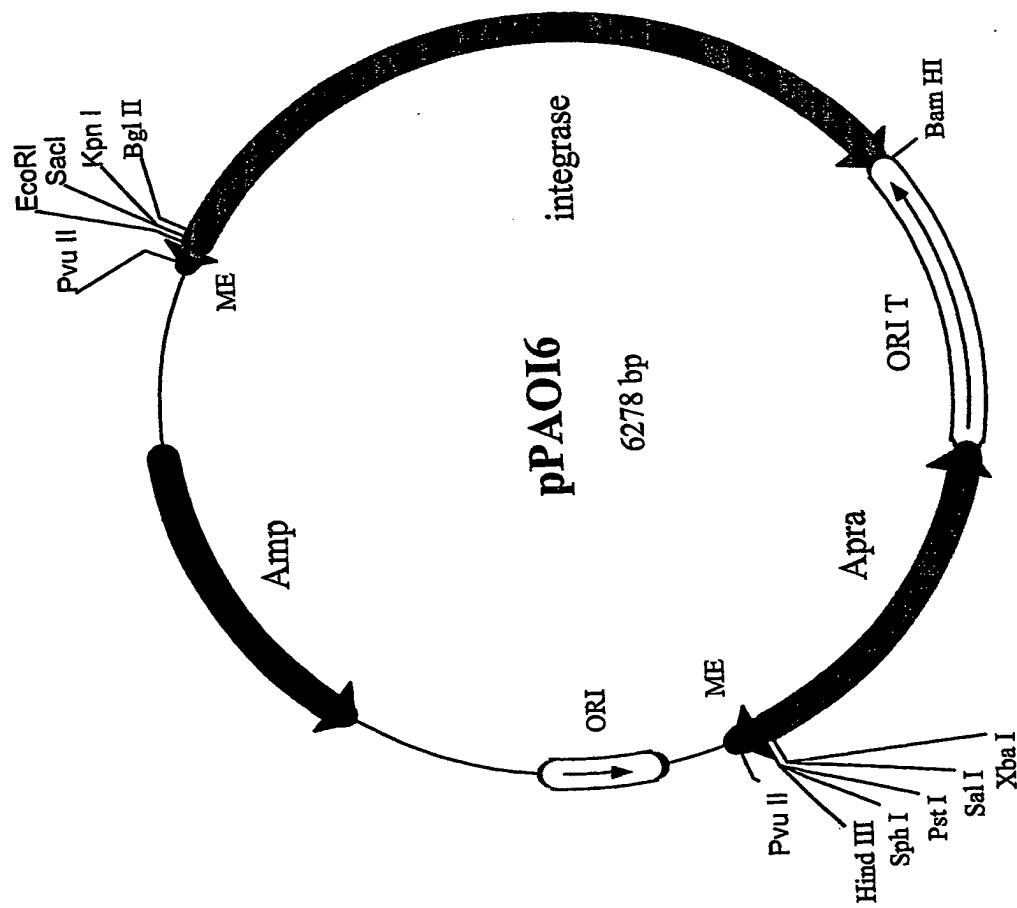


Figure 3a

6/39

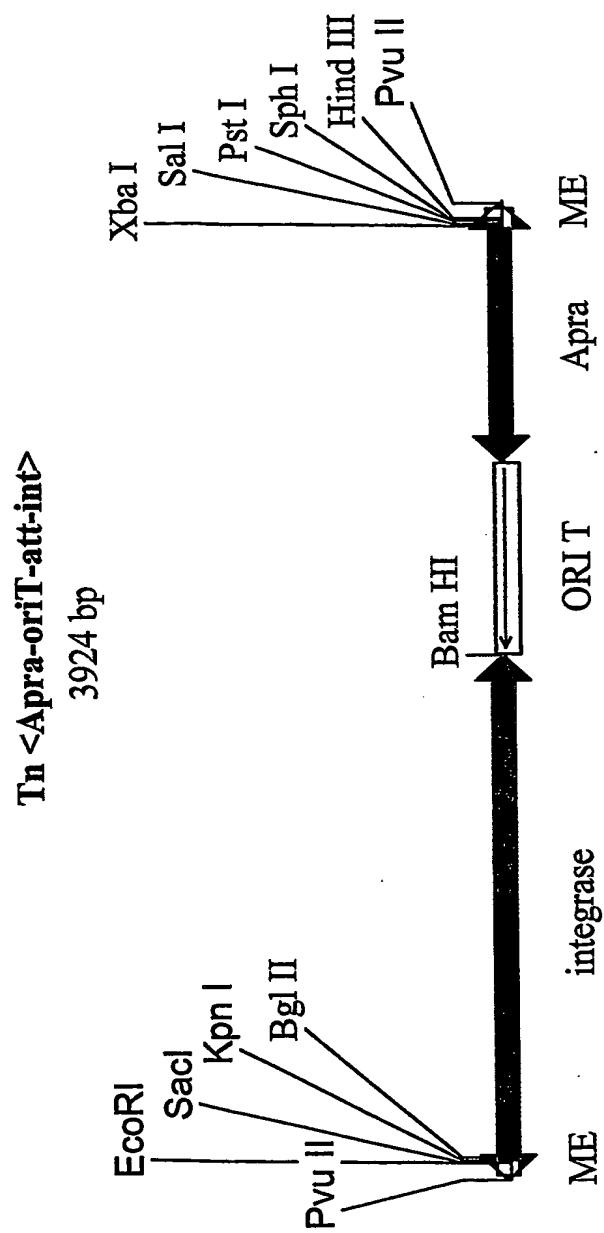


Figure 3b

7/39

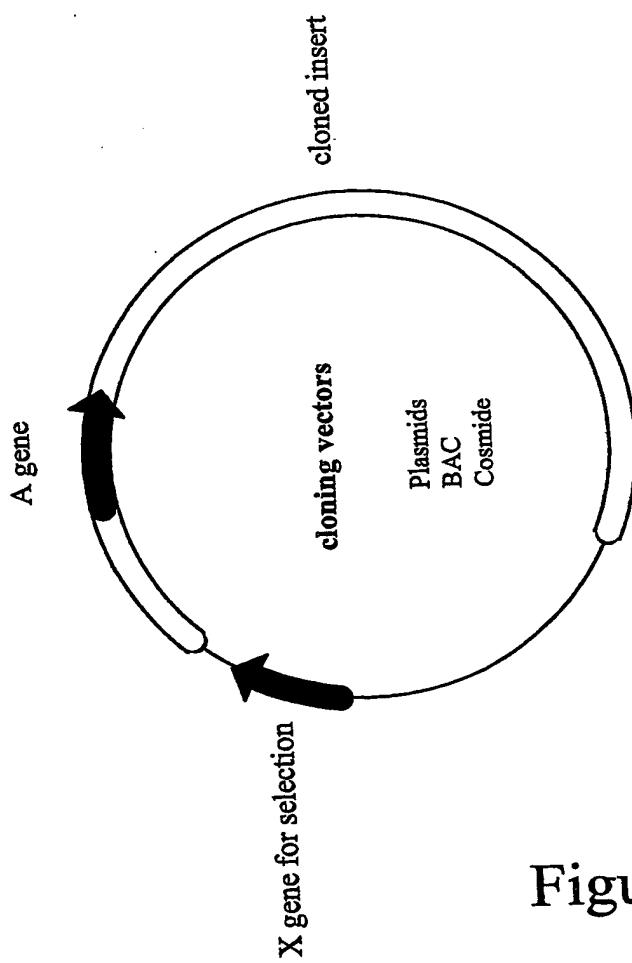
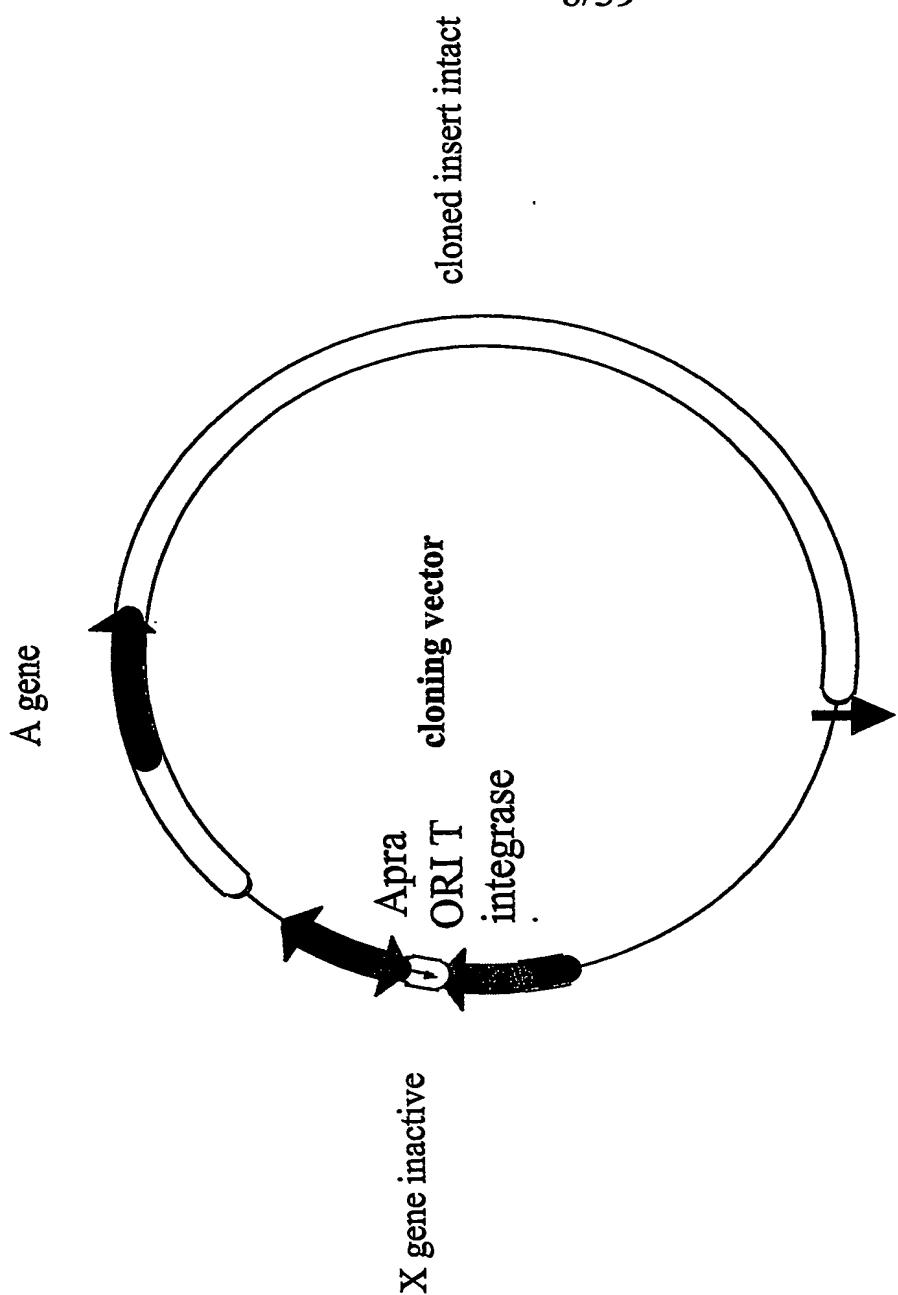


Figure 4a

8/39



Conjugative and integrative plasmid

Figure 4b

9/39

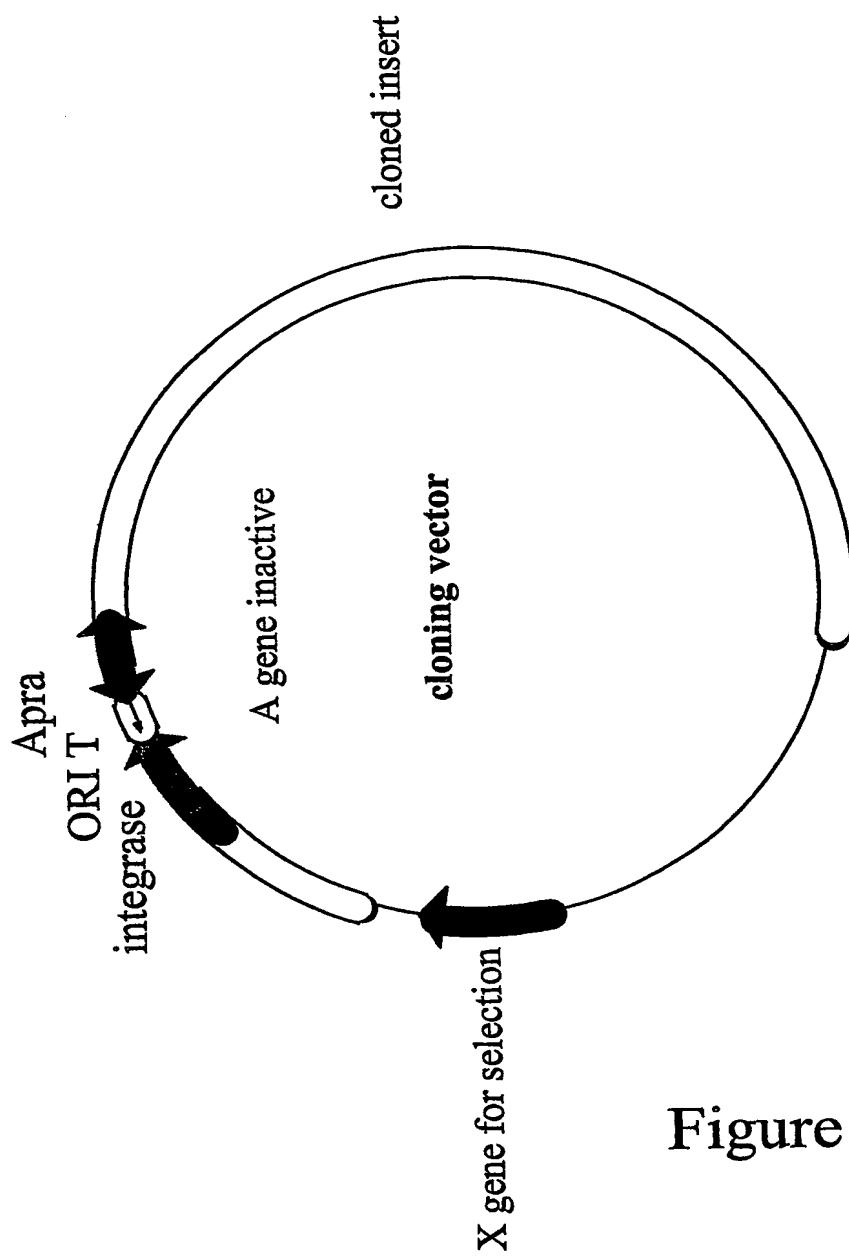


Figure 4c

Figure 5a

```

CDS 76..1134
  /note="ABC_transportr"
  /gene="TAP2 PROTEIN"
  /blastp_match="Oryzias latipes"
  /blast_score= 0.002
CDS complement(1096..2430)
  /note="none"
  /blastp_match="Anabaena sp"
  /gene="ALR1117 PROTEIN"
  /blast_score=2e-18
CDS 1178..1624
  /note="Gram_pos_anchor"
  /gene="CELL WALL SURFACE ANCHOR"
  /blast_score=1e-04
  /blastp_match="Streptococcus pneumoniae"
CDS complement(2506..3567)
  /note="CONSERVED"
  /gene="HYPOTHETICAL PROTEIN"
  /blast_score=0.019
  /blastp_match="Deinococcus radiodurans"
CDS complement(2906..4222)
  /note="glycosyl transferase"
  /gene="lipopolysaccharide"
  /blast_score=2e-23
CDS complement(4092..5321)
  /note="glycosyl transferase"
  /gene="glycosyl transferase"
  /blast_score=1e-15
CDS complement(6337..8502)
  /note="PUTATIVE"
  /gene="GLUTAMINE AMIDOTRANSFERASE"
  /blastp_match="Bordetella bronchiseptica"
  /blast_score=1e-16
CDS complement(8181..9530)
  /note="none"
  /gene="MEMBRANE PROTEIN"
  /blast_score=0.035
CDS complement(9531..10721)
  /note="NOEC Transmembrane"
  /gene="NODULATION PROTEIN"
  /blastp_match="Azorhizobium caulinodans"
  /blast_score=3e-07
CDS complement(10504..11274)
  /note="PUTATIVE"
  /gene="HYDROLASE"
  /blastp_match="Streptomyces coelicolor"
  /blast_score=4e-14
CDS 12874..13689
  /note="HYPOTHETICAL Meth-transf"
  /gene="PROTEIN PA1088"
  /blast_score=2e-06
  /blastp_match="Pseudomonas aeruginosa"
CDS 14195..15976
  /note="PUTATIVE, Glyco_transf"
  /note=""
  /gene="LIPOPOLYSACCHARIDE BIOSYNTHESIS"
  /blast_score=2e-06
  /blastp_match="Vibrio cholerae"
CDS 15427..16512
  /note="PATHWAY: INNER CORE LIPOPOLYSACCHARIDE
BIOSYNTHESIS"
  /gene="PHOSPHOHEPTOSE ISOMERASE "
  /blast_score=3e-17
  /blastp_match="Helicobacter pylori"
CDS 15579..16253

```

```

/note="none"
/gene="PHOSPHOHEPTOSE ISOMERASE"
/blast_score=2e-22
/blastp_match="Neisseria meningitidis"
complement(16505..17656)
/note="BIOSYNTHESIS PUTATIVE"
/gene="LIPOPOLYSACCHARIDE"
/blast_score=2e-17
/blastp_match="Thermotoga maritima"
/pfam_match="Glycos_transf_1"
complement(17657..18697)
/note="none"
/gene="ALR3073 PROTEIN"
/blast_score=6e-27
/blastp_match="Anabaena sp"
complement(18615..19304)
/note="none"
/gene="ALR4487 PROTEIN"
/blast_score=8e-07
/blastp_match="Anabaena sp"
complement(19301..20596)
/note="ATP_GTP_A"
/gene="ABC TRANSPORTER"
/blast_score=3e-61
/blastp_match="Synechocystis sp"
complement(20535..21476)
/note="PERMEASE COMPONENT"
/gene="POLYSACCHARIDE ABC TRANSPORTER"
/blast_score=6e-41
/blastp_match="Clostridium acetobutylicum"
complement(22025..22951)
/note="involved in the synthesis of a polysaccharide
capsule ?"
/gene="32.3 KDA PROTEIN"
/blast_score=3e-17
/blastp_match="Sphingomonas sp"
23155..26523
/note="peptide syntase"
/gene="mcyA, mcyB and mcyC"
/blastp_match="Microcystis aeruginosa"
/blast_score=0.0
26409..34433
/note="polyketide syntase et peptide syntase"
/gene="mcyD, mcyE, mcyF and mcyG"
/blastp_match="Microcystis aeruginosa"
34418..37500
/note="CYSTATIN"
/gene="PEPTIDE SYNTHETASE"
/blast_score=0.0
/blastp_match="Anabaena sp"
35359..37500
/note="gene cluster"
/gene="nostopeptolide biosynthetic"
/blast_score=2e-41
/blastp_match="Nostoc sp"
Sequence 37500 BP; 6199 A; 12698 C; 12769 G; 5834 T; 0 other;

```

gatcgggtcg	ggccctttcg	ctattacgcc	agctggcgaa	agggggatgt	gctgcaaggc	60
gattaaagg	ggtaacgcca	gggttttccc	agtacgcacg	ttgtaaaacg	acggccagtg	120
aatttgaata	cgactca	tagggcgaat	tcgagctcg	tacccgggga	tcccacgtac	180
cacgagctca	tctgaaagag	cccgctccag	acaccggagg	aatggaaacgc	cggcgtcgt	240
agaccgtaca	cgcagaagcg	tctcgtcgcc	ttcggtgact	ggcttgcgac	gacgatgaag	300
ctggatctca	ccagggatgtt	cgccgcagggc	agctcgatgg	gcggatcggg	cgcgatcatg	360
ctcgcgattc	gctatcccg	acgattcgcg	tggaccgtgt	cgtgggtcgg	cgtccacgt	420
cccgccgact	ctcccaatt	cacatcgctg	tacgagctgg	tgtacggccg	gccccgacttg	480
aagggtgcgt	tcgagaacgg	cacgcgggtc	tggggatcatt	ttagtgcacgt	ctggatcacgt	540
cggcagcgtc	cggaggcagga	catcggttc	atcagcttct	cgaacggcaa	gaacgactcg	600
cgatcggtct	ggcccccaggc	cgtcgaaattc	ctgaagacgc	tgcaggagac	caggcaaccg	660
cacctgttcg	tctggggcca	ggagggacac	ggccagcgcg	cgaagatgcc	ggaagggtggc	720

ggcgaacggg	agatgccgt	cgacctcaga	acgaaccaga	gcctgcggc	gttcagccgt	780
tgcacgctcg	acgacgatcc	gggagacggg	tcggattcga	gcccgtccc	ggccggcag	840
atcaactctt	acgtcacctg	gcaagcctgac	tcgatcgtcg	acgaaccgg	ccgatggagc	900
gtcgtcatca	agctggcgaa	tcgaccggc	cggtcgccc	ccgcagtgcg	cgtaacgccc	960
ggcggcgtgc	acgaggtaaa	gctcaagccc	ggcgatcagg	tgacatggac	gaactcccg	1020
ggcaacaccc	tcgtgcagcg	cggcgaagcg	gtcgcggacc	gttgggtct	cgtcacgctc	1080
cccgagatgc	aggtctcaac	gggtggaaac	aggatcctgg	tgacgaggaa	ctaggcggc	1140
ggcgtctcg	acgagccggc	gcatcgccg	cagttatgt	gcccacaaa	gctggcttag	1200
gtccgcgcgc	gcggcgtatgt	cccgtcgcc	ctgcggcggc	tttgcagca	aatccagcac	1260
agccgccccg	aacgcctcg	ctgattccgc	cacggagcac	gccggccgt	cgatgggtgg	1320
aagtccccct	gcaacggccg	gggtcacgac	gttcgggac	cccgccgca	tcgcttcgag	1380
cactttgttc	tgcacgcctc	tcgcgcgtcg	aatcgccg	accgcgtatcg	ccgatcgcca	1440
gagatactcg	cgtaccccg	ggacccgcgc	ggtcacttca	atcgtggat	ccggccgag	1500
tgcccgcacc	gcggcgtcg	gttccatgc	cacgagatc	agccgcgt	ccggccgggc	1560
gcaacccacg	gacggccaga	cctctcgcc	gaaccagagc	gcccgcgt	cggtgggtc	1620
gtagccgaaac	acgcccaga	agacgaccgg	cggttcggc	gccgggtggc	ctttcgccgc	1680
aaacatctcg	acgtcgatgc	cgttcggc	gaccgacg	gtgaggccgc	ccgtgacttc	1740
ttcgagcagc	aggcgctcg	gtcgcgtac	gaccgtggc	gccgcccgc	gccgcacatgc	1800
gactttctcg	aagcgagca	ageggccggc	ctcgccgcgg	aacagccagc	ctttcgccca	1860
cgccgcgcgtc	gacgcacgc	cgccgcattt	ctcggagatcg	acgtccacca	tgtccaggac	1920
gaaggggatc	ccgcccagg	tcggctcgca	ggcgtagcg	gccatcccgc	tgcagtagc	1980
gagcaccgcg	tcggggccgc	tcgcccgaat	ccgtcgatcg	agcacctgg	agatctccgg	2040
cgaatgcagc	aggacgtcg	tgaggggtcg	atcgcccg	agccgcac	cgctcgatcg	2100
gagattcccg	ccgcggccg	cgccgcacgc	gtcggcgtcg	gccgtcacgc	ccggcgacatc	2160
cgctcgatgc	gaccatcc	cgtcgatcg	cgccgcgt	acggatgtca	cgtcggccga	2220
tcgtcgagc	gtgtaatga	gatgaaacgc	gcaatgc	tcgccccgt	tcggggcgta	2280
ggccagccgg	tgtgtgagca	gagaggatct	cacgtcg	catccgatcg	ctctcgatcg	2340
acgcggcccg	gtacggccgc	ggcgatc	cgccggggaa	tcggccgg	cacgaccgac	2400
cccgccgca	tcacgaaatc	cgaccgc	tcagccat	ccacggccgc	gctgcccgc	2460
cagggtccgg	cgccgatcc	caccatccgc	ggtgtcc	gtctcgtac	gatcgccgc	2520
gtgaggtcgc	tcgtggcg	cgtgtccgg	ccgttggg	tgtggacg	ggccgcgacc	2580
aggacatcgc	gttccagg	gaccgc	agatggc	cggtcc	gtacacatgc	2640
tcgtcgat	tcgcggcgt	ctgcgaaat	atggtgc	aggcaccgt	gaccgacgg	2700
tcgcagtgt	caagcacgc	ggcgaggaa	gccgcgc	gttattgtc	gatgacg	2760
gggatgagcg	agagccactg	cgtcgacc	tcgagg	ggtccgg	gaggaagg	2820
cgccgcacgg	agtaggagag	caaggccggc	aggacg	ccagcgc	gccacgg	2880
agtcccttgc	ccgcgtc	gaggctc	cccact	cgccgc	cgccgcgacc	2940
gggctccgt	tcgtgtgac	agcgatgt	agagc	cacttgc	atccgcgc	3000
cgaacgataa	ctcttgc	atccggc	gcccgc	cgacgggt	cgccggcc	3060
cgggatttgc	catcgctc	tcgatcg	ccgcgagg	tgacgccc	ccgggtcgaa	3120
tgaccagg	gtgcgtcg	ttcgacac	gctcgcc	gccccggc	cggtcg	3180
cgatgggtgt	ctcgaaac	atggctc	gcacgc	ggcgcc	tcgtatcg	3240
acgactgcac	gacgggtcc	agcccgtat	ggacgc	gacgtcg	gtgtggcc	3300
ggaaacgcca	gttccgg	ccgagctcg	ggcgcc	ttctcg	ggacgc	3360
ggctgccc	gccggcgt	acgagac	ggcgcg	cttcgc	ccgggtcgaa	3420
aggcttcgat	gagcaatcg	atcg	cgccgt	gcccgc	cgccgcgt	3480
cgaagtcgt	ggcc	cccgtc	cacgc	ggcc	cgccgc	3540
gtcgaaagc	atgggtcg	atgcatt	ggatcg	gaccgc	ggccgtcc	3600
cttggcc	gagtc	ccaatcg	cgac	aatgac	gggaag	3660
cgagaatcc	gcgtc	gggtat	cggtcg	gcgaaac	tggccgt	3720
atccatggc	cgtggcc	gaaatc	gctcg	tcccagg	agggcg	3780
gatcggtt	gtactcg	gcgtgc	tgtcg	catat	acgacgc	3840
ggagcttgc	ccagatcc	gtgtcg	aatgctc	ccggac	ctcgac	3900
tgccggcc	tttcgc	tcgtccat	cgaa	ccgc	tcgccc	3960
agcagacgg	gatcg	ttggac	cgcc	cg	gtcgat	4020
tctccgt	gccggcc	ccccgc	cg	gac	gacgc	4080
cgatccc	cctagg	ggctcg	gcat	ccgg	cgatcg	4140
gcccggc	tatcg	gcccgc	acgac	gac	cgacgt	4200
actgc	gccgg	gatcat	gtcg	cg	gacgat	4260
agcagg	cggt	aatcc	ccgc	gac	gacgc	4320
ggaatcg	cgac	cac	ggatcg	ccgg	gacgc	4380
atggc	ggt	accac	accg	gat	ccatccc	4440
aaaaacacgg	cgca	tcgggat	ccctgg	ccgt	cgccgg	4500
gacaggatc	ggcc	ccggc	cgct	cg	ggccgg	4560
tcgagg	tatcg	acccgg	tatc	cc	ccatagg	4620
ccgctcg	gcc	cgatcc	cg	cc	tcgttac	4680
gcccgtcg	ccc	gttcc	cg	cc	gcaagatt	4740
ccgggtct	ggac	acccgg	gac	cc	ctcatatcg	4800
ggcgtcaac	acgc	ccac	gac	cc	cgcaac	4860

gccacgttca	cgtcgggcag	cgacagcact	cgcgcgatc	cgcgcctctgc	cacggcacgc	4920
gagtcgtcgg	tcgaccgggtt	gtccgcacg	acgatctcg	accagccagc	cggcgccctcg	4980
gtgcgtatgaa	tcgacgcccag	gcaacggctg	aggtgcgccc	cgcgcgttctt	gacggggatg	5040
acgaacgaca	cgcgcggcgc	cacggcaatt	ttcgcgtca	tcgatttcg	agcgcggccg	5100
gctttcgtag	cgcggggccgg	gttggggggcc	gcgagaactc	gaacgcgccc	agatccggcg	5160
ccccctgtct	cggccgcga	tcgaaatcga	gccccgcctg	cggaatctcg	atgcccgggt	5220
cgatcgccctc	cctggcggt	gccgacagg	gcagatcgca	cgaggcccg	tttgcgaacc	5280
actctgcggc	ggcggtcg	acgttgc	tcgcctccgc	ctcaccgccc	ttacggcgca	5340
gatctgccc	gctcg	cagg	tttgtgcgga	tgagccgtt	cgtcacggga	5400
cgatcggttca	cggcgtcgcc	gtgccgagcg	tgaagacgct	gttgcgtcg	atccggaa	5460
cggtggacgc	gttggcttcg	atcgcttcgt	cggccatttc	gttcagggtt	cagatcacgt	5520
tgttccgcac	gatgcccgc	tgtgtatcg	tctgtcttc	gcccgtcg	gcgtaggccg	5580
tgatccccgc	ggccaggccg	atggcgatgc	ccctgaacga	atgcacgatg	acgttccgtt	5640
cgatgagcgt	gtcctgcgag	ttcggccagg	cgaggatcg	cggggccgt	cgcgcgcgc	5700
ccgattgcgg	gcccgaatc	cgcgaggaa	cgttgcgcg	gatgacc	ccccggggccg	5760
ccaggatgtc	cacgcgtcg	gtttaatcg	acggcgcgt	cgtcgtgtac	tcgaagcg	5820
agcacgcac	gaggcatacg	tccgcgaagc	gtccgtcgt	gcccgcgt	cctttag	5880
attgtgtcc	cgcacatcgat	agctggacgt	tgtgcaccgt	ggctcgggac	gcgc	5940
cgccgcgcac	ctgaatcg	tgttacc	cgtggccgac	cgtcagatcc	gcaatcg	6000
cgtcagtggc	gccaacc	gagc	cgatcg	gccggt	ccgcgcgc	6060
gaatgacgac	gcccgtcc	tcgcgg	ggttccgg	cacgg	gggacgtt	6120
tgctggccat	cggcgtcg	cgttactc	cgttctcg	cagaatcg	gtgcgcgg	6180
tgacgcgtt	gagcgcct	acgagctcg	acgtccgcg	cacgg	ac	6240
tcgcccggc	cgc	ccatcc	cagaatcg	gcacgg	ccggacgg	6300
gaaggc	gagtgc	gcccac	ggc	cgact	tgcacgt	6360
gtgttgcgc	aacc	agag	cgaggat	gaggat	aggat	6420
tctcccg	cg	gtgtcg	cgat	gatgg	gcaacat	6480
cgtgcgc	cc	cggc	acag	ctcg	gtc	6540
ccagcgtc	aagt	ctgt	gat	gtac	ccagg	6600
gttgcac	tcg	acac	tct	cccg	ggc	6660
cgagttcc	ac	ttt	gag	cc	ccgt	6720
cgtgcgtc	cg	ccat	gc	gatc	ccat	6780
gaacggc	cg	cac	gtc	cg	tg	6840
ccgatgt	tc	ac	gat	at	ccgt	6900
gtcgagg	gc	gtc	gag	cc	ccgt	6960
tgtgaag	gt	gtc	ccag	gt	gt	7020
gtcgat	ga	tc	g	tc	cc	7080
cgccagg	gt	ctt	gg	cc	gg	7140
gtcccgc	aa	atc	atc	gt	ccat	7200
ggtgag	acc	atc	cg	cc	gt	7260
ctcgaaga	cg	gt	gg	cc	cc	7320
ggcgc	gc	gac	cc	gt	cc	7380
cagcac	cg	gt	gt	cc	cc	7440
gagc	tc	gac	cc	gt	cc	7500
cg	tc	cc	cc	cc	cc	7560
cg	tc	cc	cc	cc	cc	7620
cagg	tc	cc	cc	cc	cc	7680
gag	tc	cc	cc	cc	cc	7740
tcc	cc	cc	cc	cc	cc	7800
cact	cg	cc	cc	cc	cc	7860
cat	gg	cc	cc	cc	cc	7920
cga	ac	cc	cc	cc	cc	7980
cg	cc	cc	cc	cc	cc	8040
gac	cc	cc	cc	cc	cc	8100
cg	cc	cc	cc	cc	cc	8160
gac	cc	cc	cc	cc	cc	8220
agg	cc	cc	cc	cc	cc	8280
gg	cc	cc	cc	cc	cc	8340
gca	at	cc	cc	cc	cc	8400
ccg	cc	cc	cc	cc	cc	8460
ccg	cc	cc	cc	cc	cc	8520
ccg	cc	cc	cc	cc	cc	8580
ccg	cc	cc	cc	cc	cc	8640
ccg	cc	cc	cc	cc	cc	8700
ccg	cc	cc	cc	cc	cc	8760
ccg	cc	cc	cc	cc	cc	8820
ccg	cc	cc	cc	cc	cc	8880
ccg	cc	cc	cc	cc	cc	8940
ttggcccatt	tc	tc	tc	tc	tc	9000

atgcccacgg cgacggcgtt gttgtcaccc a诶cgagtaga cgtcggtgaa gttcttgatg 9060
 ccggaaatgta acaggagctg a诶cccagccc tgcttggcg cctcgaaccc gagggagagc 9120
 gagatcacga ggaagacgag cgcacccgc ttgacgtccg tcgtgaggac tgcgagcagg 9180
 tacgtcatca cggtgactt catgaagtgc atcatgtac cccacgcgaa gtcctggtag 9240
 ggcgcaccca gcgtcgacga caggctgtgg a诶cccgaaga aggcgaggag cgccagccgc 9300
 gcgtcgatcc gcagcgctg cccggacagg a诶cgatcg cgacgtgtt catgcccgc 9360
 aggagcgaga tggcatgtg gatgagaaag tcgctccaca cccagagctc cggccgaaag 9420
 tacgcgtatga acaggttagaa cagaatcgcc tagaacggcc cgccgacgcgt gttgaaggcg 9480
 ccgaaggcga ggagggcggag gacgaaagcg gctcgaagca ttgcccgtgtt ttacctgtta 9540
 tcggtcccga tggtcagag cggctgcgtc cgggttcctt ctacattgaa caccctttag 9600
 agggccgaa tcgaggtgaa catgagcacc acgaaaagaa tcgcccacac caccatgtag 9660
 gcaagaacc cccgcgtctt gttagagctt tccggattt gcacggggct gttcgccagc 9720
 atcccgatgt gcaggttaga ggcgaacatc cccgcgcga cccgcgaa caggatcagc 9780
 tcgaggtat a诶cgacgat gaagatcgcc cggcaagagg cgccgagggt ggcgtagaac 9840
 accatgctca cgagcagccg ctccctcggt tagaaggca a诶acttccg gtacgaggcc 9900
 ggcggcggat ctggccgtat gtcgggaaac tccgcgtacc gcttcgtcgc catgaagaac 9960
 ggcgcacca tccagtagca gatcacgacg gacatcgccg gacacgcgtc ctgaatcagc 10020
 gggaaaccagc cgaggaggag cccggacggcg ttgttcgcg actcgctcag cacgtcgagg 10080
 tacggccact ccttcgtcg aatgggcggc acgttgcgtc tgacgcccag caccaggagc 10140
 gccaggccg acagcgcgaat gtatctgtt acgagcagcg cgacgcacgaa tccggccacg 10200
 cccacgagga tccactcggt gttagccgc gccggcttga tcttcccgga gggcaccggc 10260
 cgggtccgtt tctccggatc cagcagatct ctcggccgt ccaggagctc gttcagcacg 10320
 tagttgtctcg a诶cgatgag gcaggctgcg gcgacgcgtc ggcgcagccg cggggacggcg 10380
 gtccaggccg agagctgcgg ttgttagaaag aacgcacccg gacacgcgcg cagcatgaaac 10440
 gcgcttcttgc accgtgtatc gatccgcgcg atctggact acggccaaat ccgcgaaccg 10500
 gagctagccg ctgacgacat cttgtgcgtc ctgcagccgc cgatgcgcg accatccgc 10560
 gccggctgtg acggccgtggc ggccggcgat gaggcgccgag gagagtccgt cggccatcgc 10620
 tccttctccg tccacgtccg gtgcgttcccc gacatacagg acttcccgag gtcggagtga 10680
 ccacatctca cacggccatcat ggaagccgcg cgggtcgccg ttcaaggcgat tcacgttgct 10740
 ggcggtcgtg cagagegcga gcgaaaaatg ctgcgcgtg ccgagccgcg ctagctgtct 10800
 ctcggccgcg tagtccgaca gcacgcgcgat cctgagcccc gccgcgcgaa gcgagtcgag 10860
 cgccgcgcg aacccgggc gccggcacca cccgcacatac ttgagccgc ggcgcacccat 10920
 ccattcggtc accgtcgccg ccacgcaccc gttgtccacg ccgagccgcgat ggcgggtcg 10980
 cgcgatctgg cgccgcgcgat gccgcgtc ggcgcgcgcg aggccgcggg gatccctatg 11040
 cgcccgccgg tactcccgca ctagcgcgcgcg cttgtccact ccgcgcgcgc accggaaacgt 11100
 cacgagcgcg gttggccgcg gtcggctgcg catggccgcg cgcacccggat cctggccgt 11160
 cagcgtgcgc tccacgtcg aacgcacccg cttgaccccg ccgacacggct tcactctgc 11220
 aggcccacccg tcccgacatat cggccacacc tttgacgcgt ccgcaagggaa cacgtcacga 11280
 acccacacccg cagccagccg cggccacacgc accgcgcgtc ccgcgcgcgc tgccgtccgc 11340
 cagctcccgcc gccaggccgcg cttgtcgatc gccggccgcg ctagcagag cagaatccgg 11400
 acgagccgcg ggtgataccg gggatgtccg aagacgatcg cgtgaatgcc gctcatgaaac 11460
 acgatcaggc tcaggagcag cagggtgcgcg cggccgttga ccggccgcgc gaaacgagatg 11520
 ccgatgacgc cgagcaccat gaccgcgcgcg tacgatccca ttagggacac cgccggccgc 11580
 aagaacacccg accgcggccg acggtagaaac ccattccga cggccggcgat gaaatcgcc 11640
 tccagccccc agaagtctgc gaaacttcagc accgcacccg gaaacgcgtc tccggatgg 11700
 gccttcatgt agtcgacgcg ctggcgctgc gcccacttct ctttcgtccg ctccgtccag 11760
 cggctccgtt ccggccgcgcg aggccgcaccc gtcgcgcgtt atccctgttgc gcccgtgagc 11820
 gagatcgctt cccacatcgat atccctggcc gtttgcgttgc agttggccat catcagttc 11880
 agcccccggc gctgtccac gaccgttgcg gacgttgcgat ggacgtgtt ggcgcacgc 11940
 cacggccgcgat ctagcgcgcgat gttccggccg aagagaagca cggccacccg cagacgcgcg 12000
 cgaagccggat tgcccgaccc gacgagccgcg cggccatgaa ggtacgcgcac gtacggccac 12060
 atgatcgacgc gaaacgcgcg gacgcgcgcg agcgcgcacgc ccgtcgccgc gcgatccac 12120
 gcccggatgc gcccgtatgg gcgatcgatc agcgtcagac atcccaagag caacaaaga 12180
 aggagtgtga tgaagagcgt ctggacccgc accacgcgcg cggagaacag cagcgaggga 12240
 tagaacgcga acgcggccgc ggcacatgggg ccggccgtg cgttgcgttgc cctgcgcgc 12300
 atgagatatac cgagccagac accgcgcagg ctcaacggaa cctgcgcgc cccggacccgc 12360
 gtcagcgatt ccgagccgtt gaccgcgcgcg acggccgcgcg cgaacccggg aaagagccgc 12420
 gcccggatgg acgtcggttc cccggccccc caccgcaccc cgttcccgcc gacgtgtt 12480
 ctggcaagcg tcggtaggtt ctgcgtcgat cgttgcgttgc cgttgcgttgc cttgaacccc 12540
 caaatcgcc cggccggat gaccggccgc acggccgttgc acgttgcggat tgcccgccgc 12600
 tcccaatggg tgcgtcgcc ctcgtcgccgc cccggccgcgat cgttgcgttgc gggggccgc 12660
 gatggggatgt gggggccgtc catcattcgat taaaacaagc agttacggccaaacctgtgc 12720
 caccggggaa tcagcaaaaat cttgcaccc ggggtccgcgc gggccgcgc cccggccgc 12780
 ttttccggcc ggaagtcggc ccccaaggccg accccggctgg gcacagccca gtagaaatct 12840
 cattctggat tgactttgtt ccatgatccca tgaccatttc gaggggccctt cgtgacgc 12900
 tggggatggc ataagcttgc cacctgcgtat ggcgcacgttgc cgttgcgttgc cggccgttgc 12960
 gagctccaga aggctatcg ggcacccgg gtcggccgttgc aatacgccgc gggccgttgc 13020
 cagtcaccgc tggggccgc cttccattcg cgtcgcacatcg ggggtccgcg agagctcg 13080
 cgcgcgcagg gcatccggcg cggccggat gtcgcacccg gtcgcacccg gtcgcacccg 13140

gacatcgccg	cgctcctcg	ccgcgtgacg	ctcgtcgacg	cgagcgcaca	gatgttac	13200
gaagcgccg	agcggctcg	cgacggacg	ctcgccgc	gggcaatct	cgtcaggcc	13260
gacgcgttc	gtctccgct	cagcggccg	ttcgatctcg	tctacaccc	caggctcg	13320
cgccatttc	agcgcggca	tcggctcg	ctctaccgc	agattgcac	gattctcg	13380
cccgccgct	ggctcgctt	cgatcggt	aatagatgg	tgtctcg	gcttcgcgc	13440
aacgcgaagc	ccggcgagta	cgagcaactac	gatgcgtc	tcacacccgc	cgggctacgc	13500
gacgagctgg	ccgaggccg	gttacatctc	gtgtcgctc	ccgggtcgca	tcgcgggttc	13560
cccacgtca	tgaagtgc	gatgtatctc	gcccgcgg	caacgggt	cgcgccgc	13620
gcgatggagg	tgtcgatcg	gtctccggc	gaaaccgtcg	agtggatcg	cgtatgcgc	13680
cgccgggtgac	ctactggac	gttacctggg	atccgatca	agaagcg	tcaaggaaag	13740
tgaacgcct	ccgcacccgg	tcgcgggc	acgcgacgt	cgtgtcg	tcgcggcc	13800
agcgcagccg	tctcgcc	aaagatcg	tgctcg	gtccggccgc	gcctggatcg	13860
cccttcggc	cgccggcc	gtgctcg	cgcgccggca	cgtgacgcac	gtcttcggc	13920
gccccgttc	ctggcatctc	ctgagggc	ttggggc	ggccgttctg	ctcaccgc	13980
tcaagaaaga	ttgcgtgc	gatctgccc	gggcgtac	gcacgtggc	gtcgaagtc	14040
gcccgcctcg	cgacacgtgg	atcgccgc	gtatgttcc	cgccgcgt	cacccatcc	14100
atcccgccgt	ggacccgtac	ttgtatcg	cgtgtccgt	ctcgccgg	agattcacgc	14160
tgcttcgc	cagcacccg	tccatcg	ctgagatcg	gcccgcgc	attccgctcg	14220
tcgtcgagct	ggcggccgtc	ctgtccgg	tcgacgt	ggtgcgtt	cgcgggtgg	14280
gcgacgtcg	cgccctccagg	cgcgcgtcg	acgcctcc	tccgcaccc	aacttcatcg	14340
tctcgta	ggacgcaccc	gacatgc	cctgttac	gcggccgcac	gcccacgc	14400
tgtgttcgc	ggccggcc	ggcaaggcc	gtccgaattt	cgtgtcg	ggattggca	14460
gccccggat	catecgccg	cgcgatgc	cgtcg	cgccggct	gatcgatcg	14520
gccccggctg	ggccggctg	gcccggc	cgcgcgt	cgccggc	14580	
tcggccgtt	ccgcgcgc	tacgacgc	tgtacgt	gttggcatcc	gcccgcgt	14640
gcgagcacgc	gaggcgcgc	tcttgc	gggactcg	ggacggctgc	tcgggtggc	14700
tcagaaggcg	aacgattcg	ccgactact	ccgcacatcg	tcgcagat	agcgcacgc	14760
gccccggat	cgccgcaggc	tccatcg	gtcggcgc	cgccgt	ggatcctcat	14820
cgtcggtcg	gccccatcg	gcgacgt	gcgcgc	ggcgttctcg	ggccctcg	14880
cgcgcgtac	cccgaggcg	cgatcgact	cctgcacgt	cgccgc	tggcgctcat	14940
ccggcaccac	cccgcatca	acaccgt	cacgc	gtctgtcg	gcctgtcc	15000
gtacgactgg	atcatcaacc	tccaga	gtgcgc	gcaacatcc	tctactcgat	15060
cgccgttc	taccccg	ttctcg	gtctcg	cgccgtt	ccaggatcat	15120
cacccggct	caccccg	acggctcg	gtccac	acgaacatcc	cccgttaccc	15180
ctgtgagatc	gaggacatt	tectac	gtctcg	tacgacac	ttactcgat	15240
gttgcacgc	attacatcg	acgcac	acgacgc	gcacgt	agttcg	15300
cccgagaac	cgccaggatcg	ttgcgat	cctcggtt	aactcg	gttgcgc	15360
cgatgaggcg	tctcgac	actccatcg	ctacgt	cgccctgt	gtacttc	15420
agatcggtt	cgatcgcc	tcatcg	gtccca	cgcaacccgg	ccgaacgc	15480
acgtaccgc	gagatgtcg	gcccacatcc	ggacgt	gatctcg	acaagacgt	15540
gtcgacg	ctcg	tcatgg	gttcc	ctcatcg	gctactcg	15600
ccccgtccat	ctggcgat	cccgg	gttcgt	ggctgt	gcgactcg	15660
gacgttc	cattactcg	ggtcgat	gcccgtt	gtggccat	acagcacgc	15720
gaccagatcg	cattactcg	acaacg	gttcgt	ggctgt	tcaacgc	15780
ggcccaatc	gaccgc	ccggcg	ccggcg	gttcgt	gcccgc	15840
ttcgcgc	gtttaa	tcgat	gtcaagg	agctccat	cggacctca	15900
cggtggcc	gtcggcgc	ctccatcg	aggcg	agcgttctcg	ctcgat	15960
tgtgcgg	cgccggctcg	gcccggact	cgccg	ggccggac	aagatctcg	16020
gttgcg	gagacgc	gttgcgt	ccgtgg	ccgcggc	ctcg	16080
tcaccgc	gcccac	tacgc	agaccgt	acgcggc	tgagtgc	16140
ggagcgcgg	gcaacgt	catcg	acgcgt	gcaactcg	caacatcc	16200
gcccgc	acggcg	aaagaaacgc	atgaaacg	tcgcgtc	cgacgc	16260
ggccgc	tgaagagt	cgccgc	tgc	ggcgtt	agatc	16320
tcgaggatcc	aggaagt	cattac	atccacat	ggcga	agac	16380
gcccgt	gagctgg	tagac	acgtgc	ttcga	gaccagg	16440
gctcg	gatggatcg	gcccgg	cggtgg	gcccgg	tgctgt	16500
cggatcc	ttgc	tcgcgt	ccctcg	gtccgc	atcacgg	16560
cgccggcc	gttcc	acttac	cgccg	ggcgtt	gacatc	16620
ggaaggc	ctggcg	agcacc	cgccgt	gcca	gacatc	16680
cgccgt	gagcgc	ttggcg	accagg	caggcc	gacatc	16740
cctcg	gtcg	actcg	acgc	cgccgt	tgtc	16800
gctcg	ccggat	gttgcgt	ccctcg	gtccgc	acgaga	16860
ggacg	gcccgt	tcgagg	ccggcg	cgccgt	ccgtgc	16920
ccacg	acttcc	accgc	tgatcg	gaccgg	ttgtccc	16980
gtccgt	cgccgt	atcatc	cgatgt	gaccgg	ttgtccc	17040
ggccgg	tttc	cattac	cgatgt	gaccgg	ttgtccc	17100
ccacg	acttcc	accgc	tgatcg	gaccgg	ttgtccc	17160
gtccgt	cgccgt	atcatc	cgatgt	gaccgg	ttgtccc	17220
ggccgg	tttc	cattac	cgatgt	gaccgg	ttgtccc	17280

cggccacgcg	cgcgcgcgcg	gcctgcagcg	cgggtctcat	gaatccgtac	acgacaggat	17340
tgtgcacgtg	gaccaggggc	cgccccgacga	gacccaatcc	cggaaggcatg	cgcgcacacg	17400
cggcgactg	ttgcagccg	ccggccccgca	gcccgttcag	gtcatacatac	cgccgccccgga	17460
cgcctcccg	gtcgagcgcg	gccccccgcag	gaccctcgcc	tgtaatccat	gcccgtcaca	17520
ggacgcggcg	gcgtttggc	gcccccgaca	accggatcgc	cagcacccg	gccccccgga	17580
ccagatgtga	gacgaggatc	tcgtgcaggc	ggggggactg	attcatggtg	gcccgtcgatg	17640
cggccgggccc	gcccagctac	gttactcgcg	tcgcagggtcc	gcccgcgcgg	tagagcgagc	17700
gcagcaggct	gcaagatccc	gacgggagga	agcacacggc	gtaccgcgtg	agatcggtgc	17760
gcccgcacgg	caggcccccc	tgaccgaacc	gttccgcgc	cgcgtgaaac	tcgcccgcgt	17820
cgtactcgag	gtggccgagc	gccccggacca	gacgatggcg	atgatcgctg	gccccgcgcgat	17880
gcccgcactg	atcgctcccc	cgccgcagg	accgcctccag	cgtcctgacc	atgccttcga	17940
gcacacgtat	ggacgacttc	gacatgttg	tgtcggtcg	cctgtatgc	gtcagggtt	18000
ccgggatgaa	accgaccggg	cccttcgccc	agaggcgcag	ccagagggtcc	cagtccctcg	18060
cgcgcgtggat	gcccgtcg	aacatctcc	gcccggacac	ggccgggtt	accatgaccg	18120
acgaagcggt	cacggcggtt	cgccaggacg	gttccggggag	acagcccccc	tcgagcggcg	18180
gaagatcgag	cgtgcgcagg	acccgcgg	cctcgtcgat	gacccgcgg	cccgatgaa	18240
tgaagctcca	gtcggcgtgc	tgtcggaaaga	gatcgacctg	ccggcgaatc	ttggccggat	18300
ccacacgatc	gtcggcgtcg	agcagcgcac	cgaattcccc	tcgcccgcg	gcgtatggcgc	18360
gattgcgggc	gcccggcaagg	ccccgggttg	cctgagagat	caccgtgt	gcccgcgcga	18420
atctcgccag	gatgtctgc	gtgcggcgt	tcgatccgtc	gtcgcacgc	atcaacttcga	18480
cgggctgata	ggtctggag	agcgcggact	gcacggccct	ggcgtatgt	cgctcgccat	18540
tgaacacccg	gacgatcacc	gacacgaggc	cgggatatac	gttctggcg	gcccggcgagg	18600
gtctcgagc	ggcttcagcg	gccccggccc	tctgacttg	cgtcgcgagc	gaacagcatg	18660
acgtttctcg	aaagccacgg	agacaacggc	gcccgttca	tccgcgtccc	cagcgtctgc	18720
gatgcggggc	cgaggtagct	gacccggcg	gcccggaa	tctctgcga	gtactcgagc	18780
ggctgttgc	tgatgtggcc	cattccccc	tggccggct	gcccggcg	gaagagcacc	18840
agccggcgcc	tgttcgttag	caacggcag	agcttcgtc	caacggcg	cgaaagggt	18900
tcggcgac	cgaggctcat	ggagaggtcc	gttgcgtcc	cgccgtgga	caccgcgtg	18960
agatcgaatg	agttcagcgc	gacccggcgc	tttccgcgg	cgtaccgtc	cgctactcg	19020
gatcgctgt	agccggccgc	cttacgcct	cgtcgagga	acccgtgcac	gaagtgaccg	19080
gtccgcagc	cgaaatcggt	catcgatcg	atgcgggtc	gcaacgcac	gaccatcatca	19140
acgagcacgg	gcccgcctc	gccccggacg	tgcccaaca	gatccatgac	gccccgttg	19200
tacgtccatc	gctcgaagcc	cagccgttc	ccgatccgg	acatcttcag	tccggcctgg	19260
caacgcggcc	accagtagag	ccgcttgagc	tgggtcacga	tcatgggtt	gtccggggcg	19320
cgcacgcac	gccccggccgc	gcatgcagcg	acccggagaa	ctccacatcg	gcccacggcg	19380
ccacacgtcg	atgttccctg	accgcacgc	tccgaactg	ccgcgacggc	acgagaaacg	19440
tctgggtcgg	gttgcggaaac	acgtgcaccc	acaggtagt	gttccggcg	acgagatgcg	19500
cgcggaaat	gaggtcgacg	gagaactcg	ccatgaacgc	ggacgcgtcc	ccgagatccc	19560
gccccgttgc	gttgcgtcg	tagacgacg	gctgtcggt	ggagcgttgc	accatgaa	19620
cgaacgtgac	gtcgtccagc	ggctcggtac	ttctcgccgt	gacccgttgc	ttcagctgct	19680
tccccggcgt	cacttcgctg	tgctgtcccg	acgactcg	gcccgcgaag	agcgagccgg	19740
acacgatctc	gacggggctc	ccgacactcg	gctgtccga	tttcctggac	gcccggacgt	19800
acgcgtcgag	gacttcgat	gcccgtccgg	ccgccttcac	cgacccggcc	agatacaccg	19860
catgcctcgca	gagggtcg	accgcctgca	ggttgcgca	gacgaggacg	atcgcatgc	19920
cctggcgctt	gaactccctgc	atgcgatcg	agcacttctg	cttggacgc	atgtcgccga	19980
cgctcaacac	ctcgatcgat	atcagcacgt	ccggccgcag	gttccggcg	atcgagaa	20040
cgagccgcgc	gttcatgccc	gacggatgt	gcttgcacgg	cgttccacg	aagtctgt	20100
cgccggcaaa	ctcgacgatc	tcgtcgacacc	gctcgatcgat	cttggacccgt	ttcatcccc	20160
tgatcgatc	ctggaggaaac	acgttctcg	gccccgttgc	gttccggatgg	aacccggcg	20220
cgacttcgat	gacccgcgc	ccgcgcggccc	ggacgatcg	gccccgggtc	gtccggcttca	20280
ggatcttcgt	cacccatcg	gggggtgg	acttgcgc	cccggttgc	ccgatgtatgc	20340
cgagccgcctc	gccccgggttgc	acttcgaaacg	agacgtttct	gaccggccag	aactccgtt	20400
cgtccagcgc	gtcggcggttgc	ccgggttgc	acagccgc	cacgagcgc	ggcacgagg	20460
cccgccaggct	gtcatgacgc	tcgccccggc	gaaacttctt	ggagacgttgc	tcgaagacga	20520
cgggggccgt	cacgtcagac	tttctcccg	aattcgaaact	cgagacgtat	gaagaaacagc	20580
cagccgaaga	cgacgcgtac	gaccgagacg	atcgccgt	ccgcccacat	cacggacggc	20640
ggctgggttca	ggagcagcac	ggccgcggaa	ccgtcgat	tcggcgtcat	ccgggttgc	20700
tgcaggggccc	actggagctt	ccgcggccacc	atttcgaccg	ggtagaccgt	cgacgcacgc	20760
aacatccaga	cgccgtatgc	gacccctgaa	aggtatcc	cgatcggttgc	gaagagggtt	20820
gccatcgaca	ggatcagg	gatggccgt	gttgcgttgc	cgacgcgtat	gaccgcggcc	20880
ggcagccaga	ggatgttgc	tccgaccggc	acgtgtat	agatcatgag	cgcgacggc	20940
accacggacg	cgatcgacaa	gtccacggc	ctcactatcg	ccggccggagaa	cgggaaat	21000
tccctggaa	agtacacctt	cgagacgaga	ttcggttcc	ccgtcaggga	gttacggag	21060
aacggagcg	ccgaggagaa	gaagttccag	acccacaggc	cgacgcacgc	gaacaccgg	21120
tacgcgaccg	cggtgtcgat	ggccgcgcac	cgatcgat	tcacggagaa	gacagccgt	21180
tttgcgtat	gcatgaaac	ggcccaccc	aagccatca	ccgactgtt	gtagcgcac	21240
agcagggtcg	gccgcgtat	ctgcacgc	agctcgacat	ctcggcgagc	21300	
atctcactcg	aatccggccag	gatggcgcc	atcgatcg	gttgcgttgc	ccggcgtcg	21360
ccggccggat	catgagg	ggtagcgtc	gagctggcc	aagagcaact	gcacgaaact	21420

17/39

ggtcgcgcgc	cgctcgccgg	cgacgtcgcc	gtttcgctc	gacggcacgg	ggctcgac	21480
gcggacgagc	gccccgtccg	atccgttcag	ccgcacatcg	tcgtacacca	tgtagatctt	21540
gtcttcgtac	tcgttcgtca	tcacgcggcc	gtggctctgg	taccagtaca	gcacgatting	21600
ccgctcgatc	cccttcgtta	tgacgtatcg	gttgatctcg	atgggctcg	tgcggccccc	21660
cacccgtacc	agcgcgcggc	cgtggatac	gggctccag	ccggccccc	gcaggcaatt	21720
catcgccgag	tggatctgt	cgccctcg	ctggctctgg	tagtagccg	tgtagagcga	21780
cacgtaggcc	tcgcgcggc	ccaggttagat	ccgggtatg	tactcgttca	cgccgagcac	21840
ccccatgac	tccttggta	agggcgcgct	gtcgcgtcc	gtccaccgct	ccatctggaa	21900
cgggacgttc	gcccggcgt	ggcgaagcgg	gacctgtcg	gcccgcac	agcgcgcggc	21960
aaacgcgcgc	gtgcccggaa	aacacacgct	gagaatgatc	agtcgtctgg	tcatgaatct	22020
ggactcacac	cgtggccggc	gccccgtacc	ctcggggcgg	aggcgtgaag	cgccgcagga	22080
cccagaccac	gacgaacaac	atgacgaacg	cggtcaagaaa	gacgagccag	cccgagaaac	22140
tgtggaaagaa	cccttcgtcc	gccccctcgc	cgttagtgc	cgccggccac	cccgtccgg	22200
ccacacgcgc	gcccgtccgc	ataatcgca	tcgggatctg	cgagagccg	atccgcaggc	22260
gcatcgccac	gcccgtatcc	gtgaagtagc	cgtagacat	gcccagctg	agcagcgaga	22320
tgagcgacc	gatccctgc	caacgcgtcc	ccacttcgc	cgtcgtgtgc	gcccgtccgg	22380
tcacgttgc	ttcgcgcgc	accgggatgc	tgagcgcgg	cagcacccc	tcgcccacgc	22440
gagacgcgc	gagctgcgc	ggaaatgcga	tctgattgaa	gatgatcgac	gggatcggg	22500
tcatcagcac	caggaacgc	agcggaaac	cgagcaccc	gagcttgc	cagccgagca	22560
gaaacacgat	cgcgcggta	agcgaggct	gtatcgac	gcccgtcaga	aagagctcgg	22620
cgccgagcaa	ccccgcgac	agcagcgc	gcccggcag	cacgatgc	agaccgagcg	22680
cgctcgccgt	gtccggcgc	gccgagagac	gcccggcgc	ctcccaaggc	aagttagccg	22740
cgagcgttac	gatcagaatc	ccgtgggagt	agttgtcg	gttgcattcc	tcctgcgc	22800
accgcacgat	cacccgtcc	taaaggagca	ggaaactc	ggcggcgata	ccgagcgaa	22860
ccagtcgtcc	ctgaaacca	ctcatagca	tcgggacg	tatagcaaa	ggcttcgg	22920
ggcagaact	ggcggccgaa	acacgttacc	atcacgtca	caggaaat	gcccgtaaat	22980
gacgggattt	gcccgtgg	ggggccctcg	gaatctacc	aatggatgac	aaatatccct	23040
atcaactctt	cctttgact	gcgttagc	gaagatcatc	ggccttca	gatcaggac	23100
aaatgtcatc	tgtcaagaa	agtaatgtct	ctggccac	cgccggtaa	gtgaaatccc	23160
ccgacagaat	cgcgtaccac	cgcgactgg	atccggcag	aaaagtcgc	tcgtcgccc	23220
aaaaactgca	ttacaggaac	gccaggtcg	cagccgtt	gtttccgt	agtattttg	23280
ggcattgcag	gcccgttcgt	ctgccccgg	gccccgt	gcccggggc	cgatcgc	23340
tggagcgagg	caggagcagg	catggatgt	tccgtgag	agccgcgtt	gggagacggc	23400
cgggctcgcc	gggagctcg	ccacctggc	gtggaccgc	cgccgcgc	atccctgac	23460
gcccggccgc	taagtttga	cggcgtctcg	ctctcgatc	gcccgttgc	tcgcccgc	23520
aacgcgtcg	cgcacccgc	gacgcgtat	ggcgtcgatc	ctgacgtgt	gttgcgc	23580
tccctcgacc	ggtccatcg	gtcgtatcg	gggatctcg	ccgtctca	ggccggccgc	23640
gcctacgcgc	cgtcgatcc	tgcgtatcc	aaggaccgc	tcgagatc	gttcaggac	23700
gcccggccgc	ggatcg	cacccgtcg	cgcgtgag	gtctgtatcg	cgtcgtatgg	23760
gtcgagctca	tctgcgtcg	tcgggacg	ccccggggt	tggcggaa	ggccatgc	23820
ccggacgcgc	cgctcgatct	cgaccac	gctacgtca	tctacac	gggatcgacc	23880
ggggccccc	agggcgctgc	gtgcccgt	cgtccgtt	cgaatctcat	cggtgcgc	23940
gtcgacgc	cgccgtcg	cgccggcgc	cgcacgtcc	agttcactt	gccgagctt	24000
gacgtctcg	tccagaaat	tttgcgac	ctcgcgacc	gcccgcgag	cggtctcg	24060
tccgaggaca	cgccgggg	tccgcgcgc	ctgctcg	tgcgcgc	tcgatcg	24120
gcccggctgt	atctccctt	cgtcgcgtt	cacgactt	ccgtcacc	caccgac	24180
ccggatcg	cgccgttgc	tgaggtc	acggccgg	agcagctcc	ctcgacccc	24240
gcccgtcgcc	agtggttcgc	gccccatcg	cagtgcac	tccacaacca	ctacggcc	24300
gcccggccgc	atgtgtcg	tcgcacac	ctgactag	ccccggcgg	atggccgg	24360
ctggcccg	ttggcgt	gctgcccgg	gtcgagat	tgcgtcg	cgagcatcg	24420
cagccgtcg	cgcacgg	cgtggccg	atccctatt	gccccgt	cctcgatcg	24480
ggctacttga	atcagccg	gcccacgg	gaacggtt	tccgcat	gttcgc	24540
gatgtcg	tgtatcg	cgccgtatc	ggccgtgt	gcccgcgtt	cgccatcg	24600
ttcgtcg	ggacggac	ccaggtaa	attcgagg	tccgtgtt	ggacacttac	24660
acggggatgc	cctacacc	ggcgagat	cgggatgg	ccgtatg	cgccgcgc	24720
atccgtcg	atgcgcac	cgagggtc	gagatgg	gccccacgg	catgggt	24780
ttccgtcg	cgccccact	cggtgtat	agcgcac	atgtctcg	cgccgcgtc	25080
gaccacgt	ggggccac	gacggcctcg	ggccca	acgtcg	cctctcg	25140
gaggccgg	acttacc	cggtccgg	gcccgtt	atgcccgtt	gatgaa	25200
gtcgccgc	acttgc	cggtggact	ctgacgc	tcctcgcc	cgccggcc	25260
accgtcg	cgggcg	catcttcgt	ggcgcac	cgacac	cctgctcg	25320
atgttca	cggtcg	gctcag	ggatgg	gccccacgg	atggatgtt	25380
ttcgacgc	tcatgg	catcccc	atctcg	cccagg	gatccggc	25440
ggccgtt	acacgg	gacccgtt	aggta	cgacgc	gttggccgg	25500
ggccgtt	acacgg	gacccgtt	aggta	cgacgc	gttggccgg	25560

cggccggccg	ggcgccgtcc	gtccgtcgcg	ctcgactggc	gtacggagcc	ggtcgcccag	25620
cgtcagctcg	ggcaactgt	cagcgccgat	gcccggacg	cgatcgagat	caggaacgtg	25680
accaacgcgc	ggctgctca	gcccggccgc	tgcttcgatc	tcgtcaagcg	tgccgactgg	25740
acccggacgg	ccggcggagac	gcttgagagg	gcccggccgg	acacgcgcga	cgcgatcgac	25800
ccggaatcg	ggggcgct	cgcgaacgc	gcccgtacg	acgcccagtt	cacctggtcg	25860
ggcgagatcg	gcccgtcg	cttcgatcg	cggttcgagc	gcccgggcac	cgcgtcgctc	25920
gcccgtccc	tgactcgtca	cgcggttccc	gcccggagc	cgcgggcat	tcagcagtac	25980
gacgacgatc	cgctccgcga	tgcgtggcg	cggcgcctca	cgcggagct	tcggcgattc	26040
ctgcaggagc	gctgtccgga	tcacctggtg	ccgtcgccct	tcgtcgta	cgatgcgtg	26100
ccgctcacgc	caagcgggaa	gatcgatcg	cgcgccttc	cggcgcggca	gagccggcgt	26160
cccgatctgg	acgtggcg	tgccaagccc	gccacggagc	tcaacgcga	gatcgccgac	26220
gtctggcaga	cgacgctcca	gatctcgtcg	gtgggattgc	acgacaactt	tttcgatctg	26280
ggcgggact	cgctcctgt	cgcgcaggcg	ttcgagcgta	ttcggccgct	cgtgggggc	26340
aagcagtgg	cgatgatcg	gatgttccag	tacccgacgc	tctatcgct	cgccgggtt	26400
ctcagtgagc	ggccggacgc	ccgcgcgggg	cagtttccg	gcccgcggag	tcgcggacgc	26460
cgccagcgc	agatgtgt	cggtcagggg	cctggcttca	gcagagtcg	aggggtccgg	26520
taatgcagaa	cggcgcg	ctcgtcggg	ttggccggcg	cttccgggc	gcccggcgt	26580
tcgaggagta	ctggcgcaac	ctcgtcgccc	gcacggagtc	gatcacggc	ttcacaacgc	26640
aggagctggc	gacggccggg	atcagccaga	gcatctgc	gaatccgcac	tacgtccgcg	26700
cgcgcgggct	cgtccggat	ccggacctct	tcgacgcgc	gttcttcggc	atcagccagc	26760
gcaagccga	gctgtatggat	ccgcgcgcacc	ggcttctct	cgaggcgtc	tggctcgcgc	26820
tcgaacacgc	ggcgatcg	ccgcggcgt	tcccggtct	catcgccgt	tggggccggca	26880
tgagcacggg	gatgacgaac	agcacgtacc	tgtgtcgaa	tctccacagc	cacgcgggat	26940
tgctccagcc	ggaggacgt	ctgcccggc	tgctcgca	cgagaacgcac	tacctcacga	27000
gccgcgtctc	gtacaagctg	cacctcgcg	gaccgagat	gaacgtcg	accgcgtgct	27060
cgacgtcgt	cgtcgccgt	gtgcaggcg	gtcagtcgt	gtcacgttg	cagtgtatg	27120
cgggcgtcgc	cgggcggt	tccgtgtcg	tcccgcaaca	ggagggtac	gtctacgtcg	27180
aggggcgcat	cggtcgcc	gacgggact	gcccgcgtt	cgatgcga	gccccaggca	27240
cggttccat	caacggcg	gcatcg	tccgtcgaa	gctcgaggac	gcatcgccg	27300
acgggcagac	gatctacgc	gtgatccgc	gatcgccgt	gaacaacgc	ggctcgccg	27360
aggtcgttt	cgcagcccc	agcgtggac	gccaggcga	gttcgtggcc	atggcgca	27420
gcccgcgc	cgtggcgcc	cagacgatcg	actacgtc	ggcgcacggc	accggcacgc	27480
cgctcgaga	tccgatcg	gtggctgc	tcacgaaggc	cttcgcgc	ggcggccgca	27540
cgagaacgg	gttctcg	ctcgatcg	tcaagagca	tttcggcac	ctcgactcgg	27600
cgggccgg	cgcgggct	atcaaaacgg	cgctcg	ccatcaccag	cagattccgg	27660
cgacgtcgt	ctacgtcg	cccaacccg	agtgtat	cgccaacagt	ccgttcttcg	27720
tgaacgcgc	gtcgccac	tggccgc	gtggccatcc	gcccgcgc	gcccgtgag	27780
cggtcggt	cgccggacc	aacgcgcac	tcgtgtcg	ggaagcgc	cccggccgg	27840
agcgaccgg	cgccgagccc	gcccgcac	tcgtgtat	ctcggcga	acgcggccgg	27900
caactcgac	ggcggtcg	acactcg	gtcatctcg	ggcgcaccc	gacgtcagca	27960
tcgcccacgt	ggcgcacacg	ctgcgc	gcccacag	gttccgc	cgacgacgc	28020
tcgcccac	cagccgc	gatgcgatcg	ccgtgttgc	ccgcgcggac	gccaacatgc	28080
tcgcccac	cgacgc	tccggcag	gtcccgtgc	cttcatgtt	tccggccg	28140
gcacgcg	cgcga	ggccggag	tgctggcg	cgccgcagg	ttccgcga	28200
aaaccgagcg	ctgcctc	cacctgt	tccgc	cgccgcgt	gctgtttc	28260
cgatcatgc	ggacgtcg	cgggccgc	attggctcg	agtcgtc	gttggtttgc	28320
ccgcgtctt	tgtcg	tcgcgtcg	cccgctgt	gatggcg	gtgtgc	28380
ccgacgggt	catcggt	agcatggcg	agtacacgc	cgctgtc	gccccgt	28440
tgtcg	gaacgc	cgctgtgt	gtctgag	ccggcg	gagacgttc	28500
cgccgggg	gatgtcg	gtcccgt	ccgaggc	ggtcg	gcccgc	28560
atggactcg	gatcg	gtggcc	ccaggc	ggtcg	gcccgc	28620
acgcgtc	ggccttgc	cgctgtat	ccgaggcc	tgtcg	gccccct	28680
acatcg	cgcgc	tcgccc	tcgatcc	cctggcg	ttcgagcg	28740
tcctccgc	catcg	gacgc	gggtcc	cggtgtt	cgacgttc	28800
cgtggctg	cgcgc	gca	gggtcc	cggtgt	cgacgtt	28860
ccgtcagg	tttc	ccgc	gggtcc	ccgcgc	ccgttct	28920
aagt	ccgg	cgac	gggtcc	ccgcgc	ccgttct	28980
cgcacgt	tca	ccgg	gggtcc	ccgcgc	ccgttct	29040
tgcacgt	gggt	ccgg	gggtcc	ccgcgc	ccgttct	29100
acagaggc	gggc	ccgg	gggtcc	ccgcgc	ccgttct	29160
acctgt	atcg	ccgg	gggtcc	ccgcgc	ccgttct	29220
ccgatgt	tcag	ccgg	gggtcc	ccgcgc	ccgttct	29280
atcgat	cgac	ccgg	gggtcc	ccgcgc	ccgttct	29340
agatcgat	cgcgt	tttct	gggtcc	ccgcgc	ccgttct	29400
cgagtctcg	gttca	aggc	gggtcc	ccgcgc	ccgttct	29460
aggcacc	ccgcgt	gagtt	gggtcc	ccgcgc	ccgttct	29520
ttcc	ggcgt	ccgcgt	gggtcc	ccgcgc	ccgttct	29580
gcct	ggcgt	ccgcgt	gggtcc	ccgcgc	ccgttct	29640
ttcaacagca	attgc	aggc	gggtcc	ccgcgc	ccgttct	29700

19/39

tgctgcccgc gatgccgccc ccagcgtcgt ccgtcgccgc ggcgcacgcg ggactgaagt	29760
cccgcgctac gacggggccc gtctcgaaagc cgcatcgccgc cttaagccgc atcgacggca	29820
aagccagtgc ggagctgtcg cccgagcaga tggccgcgt cgccgcctgg atcgatcgat	29880
acacacggcg cacggccgga tcaagcgac gcctggcga gtacccggcg gtgcttgcgc	29940
atccacgctc ggtcgccgga ttcaaccggc tctggaaagga gatgatctac ccgatcatca	30000
ccgaccggc ggaaggatcg aaggcttggg acgtcgacgg caacgagtagt cgttgcgtcg	30060
tggccgggtt cgctcgctt ctgttcggac accgttgcgc gatctggcgc gacggccgtgg	30120
aggaacaact ccaccgggg ttcgagctcg gaccgttgcgc gccgcgtcgc ggcgaggctcg	30180
ccggcgtcg caaggaattt acggggatgg agcgcgttgg attctccaac accggctccg	30240
aagccgtct ggcggccacg cgtttgccc gaacggtcac cgggcgcgac aagatcgccg	30300
tcttcgaaagg gtcgtatcac ggtcttcctt acgaagtgtct cagccggccg ctcgtggta	30360
acggcgaatt gcggtcgcc cccgcggcgc cggcatcgc cgccgcgcg gtgagcgaag	30420
tcatctgtct cgagtaacgcg aacccgaagt cgctcgaggt catccgcgcg caccgatcg	30480
agatcgccgc gatectcgic gagcccggtgc agagccggcg gtcgacactg cagccgaaagg	30540
agttcctgca cgagggttcgc cgctgtggcg acgagatcg cggcgcgtcg atcttcgacg	30600
aagtcatcac cggcttcgcg gtgcattcccg gccgcggccca ggcgcacttc ggcgtcccg	30660
cggatcttcg cgcctacggc aaggctgtcg cgccgcgtatc gcccatcgatc atcgatcg	30720
gcaatgcgaa ttcatggac gcttcgacg gccgcgtactg ggagttacggc gacgggtcg	30780
tcccgaggt ggggttcacg tactttgcgc gcacgttcat gccgcattcccg ctcgtcg	30840
ccgcgtcgaa ggcgtcgctg acgtacatga agtcgcagg tcccgccctg cagacgcggc	30900
tggcagatgc gatcgagcgg ctgcgcgaaacg acgcgcgcgc gatctggcgc aggacaggcg	30960
cgccgtatca gattacgcg ttcagctcgat tgatgagctt gaaactcccg cacgaccaga	31020
agacggcggt gctgtgttc ttcttgatgc gcgagcgcgg gattcataatc tgggaggggcc	31080
gtcccttctt ttcaacgacg gtcacaccc agggcacta cggccgcctc ctccgcgc	31140
tcgacgagtc gctggccgag atgcaggcgg ctggcttctt cggccgcggcc gcccactccg	31200
cattgcaggt gaccgggacg gcccggcgg acgttcgtcc gttcaccgaa ggacaggcagg	31260
agatctggct cggccggccag atggggcggg cggcatcagc cgcctacaac gaagtctcg	31320
cgctcgacct ggcggggcct ctgaatcgic cggcgatcgac ggcgcgcgtc gacgaggctcg	31380
tcgcgcgtca cgaatctcgic cgcatgcgg tggccagcga accgcgtcgic ctgcgtc	31440
ttccgggcac gtcgggtcgic gttgaatggc aggacgtgtc gggattggcg gaggacgtc	31500
agcacgaggc cgtcagaagat gtgtcgagg cgccgcacgg cgtggcgatc gacgtgaac	31560
gcggccgtt gctgcgcgtg accaccatga cgctcaccccc cgagcatcactc gtctcgta	31620
tggccgcacca ccatctcgic tttgtacggc ggtcggtcc cgctcgatctt cgcgatctgg	31680
cgacgttcta ttccggccag gttcggggcga cgcgtcgacg cctcgaggcg cccatgcagg	31740
tgagccgtt cgcgtcgat gatcgcgagg ccaagcagag tcttgcggc gcccggaccc	31800
aggcgttctg gatccggcaat ttcgacaccc tccccggacgg gtcgagttt cccggccatc	31860
ggccgcgcgc gccagtaag tctgtaccgg ggcggcggtgt gtccgtgcgc tttgacgcgg	31920
cgctcgctcg cggcccaag aagctcgcc cggaaacccg caccggccctg tttacgcgc	31980
tgctcggtcg cttcgagacg ctcaatggc accagaactt cacgtacggc agttgtctcg	32040
gcatccccgt cgggtgtcgat ctctcgatcacc gactgacccg ccagcaggat ttcgtcg	32100
tcgtgcgttc tccgcgcgc gtcgtacccca cgcgcacgtt ctcggagatc ctgaacacgc	32160
tgaggacgca gtcaactcgac ctcaatggc accagaactt cacgtacggc agttgtctcg	32220
ccaagctcgic gttggcgaag caccggggcc gcgatccgat cgtctcgacg agttcacgc	32280
tcgagccgcg catcgatcgat cccgggttcg acggatctc ggcgcgttcc ctgacgggtt	32340
cgcgcgtgac atcaagcgc gacctgcacg tgaacgtcat gtagatggac ggcggctcg	32400
tggtcgaggc cgactacacg accgatctgt tcgacgaaacc gacgggtcgic cgggtgtcg	32460
atagctatcg gatctcgatc gggccgtcg tttgtccccc cggccgttcc ctccctgagcg	32520
tgccgtgtat ctccggagccg gaccggatc agttgtcact cggctggaaac gacacggccg	32580
ccgactaccg gcgatcgatc gtcatccatc aactcgatcgaa agacgcgcgc gtcgacgacc	32640
cggaggact cgcgtcgatc tttgtggccg acggcgatcg acggccgcgc ctcaaccgc	32700
gcccggccatc gttggcgtatc ggcgtcgatc agaaggccgt cggccggccgg agcagggtgg	32760
cgctgtcgatc ggaacgatcg gccgggttca tcgtctcgat gtttgcgttgc cacaaggccg	32820
gcccggccata ttttgcgtatc gaccggatc cggccggccga tcgcaagttt ttcgtcgatcg	32880
aggattccgg tgccgtcgatc gtcgtacccg attccggatc cggccggccga tgggtcgacgc	32940
ccaggacgccc cgtgtccac gtcgacccg acggccgcgc aatcgccaa gatgtcgacgc	33000
atgatctgaa ggtcgccctg tcggccgaaacg acggccgcgtatc gtcgtatctac acgttcggat	33060
cgacggggaca accgaaggggc gtcgtatcg ggcaccggca gtcgtatcg tacacgtgg	33120
cgggtatcgatc acggatcgatc ttaccggccg gtcgtatcgatc ttcgtcgatcg tcgaaatcgatc	33180
cggcccgatag cgggtgtacc gggcttcgtcg cgtcgatcgat gggctgggtt ctgcacgtcg	33240
tcacctcaga ggtcgccgacc gacggggccgg cgtcgatcgat gtttgcgtatcg ttcgtcgatcg	33300
tcgacttcctt gaagatcgcg cgtcgatcgat tcgtcgatcgat gtcgtatcgatc caccggatcg	33360
cctcgatcgatc gccggccgcgc tttgtatcgat tttgtatcgatc ttcgtcgatcgatc caccggatcg	33420
tggaaagatcgatc gcaatcgatcgatc gcaatcgatcgat gtcgtatcgatc ttcgtcgatcgatc caccggatcg	33480
agaccacggatc cggcgatcgatc acgtacttgcg cccgtatcgatc cgggtatcgatc ttcgtcgatcgatc caccggatcg	33540
cggtggccgc cggcaggccg ctcggaaacg tccggccatc gtcgtatcgatc cggccggatcgatc ttcgtcgatcgatc caccggatcg	33600
cgccggccatc gattggcgatc gtcggatcgatc tttgtatcgatc ttcgtcgatcgatc caccggatcgatc ttcgtcgatcgatc caccggatcg	33660
ggtatctcaat ttcgtcccgac ctcaccggccg acggatcgatc tttgtatcgatc ttcgtcgatcgatc caccggatcgatc ttcgtcgatcgatc caccggatcg	33720
agccggggacg ggcgtatcgatc cggaccggccg accggccggccg tttccggccggcc gacggaaaca	33780
tcgatccatc cgggtcgatcgatc gactatcagg taaaatcgatc cggatccatc gtcgtcgatcgatc	33840

20/39

gcgaaatcga	agcggcgctt	cgcgcctatg	agggcgctga	gcaggccgtc	gtcgtgtgc	33900
gaaaagatca	gcccggcgac	gagcgcttga	tcggctatgt	cacgaccggc	ggccgcgcct	33960
cgatctcgat	ggggagctt	cgcacgtcgc	tcaagcagag	gttgcggac	tacatggtgc	34020
cgccgggtat	cgtcgctgc	gacaagctc	cggtcacgtc	gattggcaag	atcgatcgac	34080
gcccctgccc	ggcgcggccc	gagcgcggca	gtttcgatc	ggagttcgtc	acgcccggca	34140
ccgagacgga	gcccggctt	ggggagatct	ggagcgcggt	gctcggttctc	gagaagattg	34200
gcccgcacga	caacttcttc	gacctggcg	ggcactcgct	gctcgccggcc	cgcaccggca	34260
tgaaggctcc	cgacacgttt	cgcgtggacg	tgtcgctcat	cacgttcttc	gaaatggcca	34320
ccatcgccat	gatggggca	gcaatcgacg	ccggtaccgc	ggcccttgcc	tcccgacca	34380
tcacgcccag	ggcccggtc	agggcccgca	gggatgagct	gttccggccca	tgaaccccg	34440
ccaccccgac	actgagttcc	ccgacactcg	gcccgcacgg	atcgccgcgg	aagcggggcga	34500
cgttttctg	gcccggctg	cgtacggcca	ggggcgctt	tggttccctgg	cgcagctcga	34560
gcctgaaagc	tccgcctaca	acattccat	cgcgctcaga	ctgaaggggc	gcctcgaccc	34620
cgccgcgcctc	gaacgcgc	taacacat	cgtccagcgt	cacgagacgc	tgccggacac	34680
gttctcgctg	caggcgatg	agctcaagca	ggtcatctt	ccgacgtcgc	gtattccgt	34740
ctccctcg	gacatcgctc	tgcctcgtt	tgagaaacgc	gaggccggat	ggccgcgcgt	34800
tgccggccac	gacggcgcc	gcccgttgc	tctggcttag	ggcccgctgc	tcaaggctgt	34860
actcgatcatc	ctcgagcc	ccgatattcg	cctcctgtatc	accgtgcata	acatcgatctt	34920
cgacgggtgg	tccggggga	tcttcatccg	cgagctggcc	gaggcgatc	cggtcgagct	34980
cgacgggcgc	gacccgaagc	tgccggagct	gccgatccag	tacgcccgc	ttgcccgcgtt	35040
ccagaaggaa	cagacccaa	ggccggccct	gccgggtctg	ttgaagact	ggacgcagga	35100
gctggcgggg	gcgtcgatga	cgctcgac	gccgacggat	cgtccgcgtc	cgccctgca	35160
gaccccccgg	ggccgcctcg	cctcgatcg	gctcgacgg	gcgtatcg	cgccgcgtcgc	35220
cgcattcagc	cgggggcgaga	acgcgacgg	gttcatcg	cttctctccg	ccttcggcgc	35280
cgtgtcgat	cgtgttccg	gacaggaaga	catccctatc	ggcagcccg	ggccggcccg	35340
cgatcgatcg	gagcttgcgg	gcctgtatcg	gttcttcatc	aacacgtt	tcttgcggct	35400
caacgcgtcg	ggagatccca	cgtttgcg	gttgcgttgc	cgggcgccgc	gcccgtccg	35460
tgcggccatc	gcgcattcagg	acctggcg	cgatcggtc	gtcgcagacgc	tcaacccg	35520
gcccggccgg	gatcgatc	cgatctatca	agtatgtt	gcccggatc	cgccggcg	35580
gaagatgacc	tccggggga	tgcgcgttgc	gcccgttgc	ttccacaacc	gcagcaccaa	35640
gtgcgcac	acgcgtcgagg	ttgcgcgaa	cgccgcacgc	gtcaccctgt	acgtcgagta	35700
cagcgatcgac	ctgttcgat	cggcgcacgg	ggcgtcg	ctgcagcaat	atcgccgt	35760
gcttcagacc	gcgcacgc	acccggacca	gcccgtctcc	gagctcg	tgcttgcga	35820
cgacgacgc	atgcagatcg	tgcacttgg	gaccgg	gcccgtct	acccgcgc	35880
cgtcgtctg	gccacggct	tcaagagacg	ggtccgcgc	actccggc	cagtggcgc	35940
gacgttggag	ggacagcacc	tcacgtacgc	ggatctca	cgcgcaggc	accgcctgg	36000
ccggcggtt	aaaggcgatcg	gtgtggggcc	gaacgtcg	gtccggcg	cgctcgacc	36060
gtcgttcgat	ctgtcggtt	cgctcctcg	cgtggcg	gcccggcg	cgtatctcc	36120
gctcgatttc	gctcccccgc	aaagacgg	tgcgttcat	ctcgccgaca	cgcgcgc	36180
catcgatcg	acgctcg	gatggcg	ggcgtcg	tctttcccc	tcttgcgt	36240
ggacgacgag	ctgcccggca	ccgcccgg	gcccggac	gatctcg	cgcagtcgac	36300
cgagacccac	ctggccat	tgtatgtac	ctcggcg	actgg	ccaaaggcg	36360
cgcgatcac	catcgccgg	tggtgcgt	cgtacgc	acagactac	tccgtcg	36420
cgacgtcg	cgatcc	acatcg	gacgtcg	gacgcgtt	cgatcg	36480
gtggggcg	ttgctgaa	cgccgcgg	ggtgcgt	cccgg	tcgtcg	36540
ggagacgctc	gagcgac	tccggcg	cgagg	tcgtcg	tcacgc	36600
gctcttcaac	gcccgtatcg	acgcgaa	cgccatt	gtccgg	tcgac	36660
gatcgccgc	gaa	cgatcg	cgtgg	gccc	atcgcc	36720
cacgcggc	gtgac	acgg	cgatgc	cgat	cgcc	36780
gatccgc	acgc	tcatt	gtccatt	atcg	ccaa	36840
tacgcagg	ctat	atcg	acacgg	cgac	cgat	36900
gctctgc	ggcg	ggct	ggcg	ccgtt	ccgc	36960
ggaacgg	gtg	cgg	ccat	ccgc	cgat	37020
agac	ccat	cgat	cgat	cgat	cgat	37080
gatcaa	act	cgat	cgat	cgat	cgat	37140
gggagacgtt	ctgg	acgg	ggat	ccgg	ccat	37200
cgcctac	gtgc	cgcc	atgg	ccat	cgat	37260
gaaggcg	ctgg	cgat	ggat	ccat	cgat	37320
gcgaacgg	gccc	ggca	cgat	ccat	cgat	37380
cgcgc	ccgc	gcgt	ggat	ccat	cgat	37440
ctgg	cgaa	gtgc	ccat	ccat	cgat	37500

//

21/39

Figure 5b

```

CDS complement(3..914)
/blast_score=2e-66
/blastp_match="AE004644.PA2177 Pseudomonas aeruginosa"
/gene="regulator hybrid"
/note="probable similarity to prokaryote sensory
transduction proteins"
/product="sensor/response regulator hybrid"
CDS 924..2168
/note="none"
/gene="ligase"
/blastp_match="AP003013.MLR8297 Mesorhizobium loti"
/blast_score=e-152
/product="2-amino-3-ketobutyrate CoA ligase"
CDS 2207..3190
/blast_score=e-151
/blastp_match="AE008872.TDH Salmonella typhimurium"
/note="none"
/gene="dehydrogenase"
/product="threonine 3-dehydrogenase"
/pfam_match="PF00107; adh_zinc; 1"
CDS 3373..4455
/note="putative"
/gene="methyltransferase"
/blastp_match="AE001866.DR0026 Deinococcus radiodurans"
/blast_score=1e-08
CDS 4546..4959
/blast_score=2e-11
/blastp_match="AP002997.MLL1617 Mesorhizobium loti"
/gene="unknown"
/note="pfam00263, GSPII_III, Bacterial type II and III
secretion system protein, Expect = 7.8"
CDS 5176..6192
/blast_score=5e-98
/blastp_match="AF064070.PE20 Burkholderia pseudomallei"
/gene="glucose epimerase"
/note="putative "
/product="UDP-glucose 4-epimerase"
/pfam_match="PF01370; Epimerase; 1"
CDS 6331..14043
/note="substrat AT, malonyl ; zinc depend dehydrogenase ;
zinc dependent adenosine deaminase putative"
/gene="PKS I"
/blastp_match="AF285636.WCBR 2547 Burkholderia mallei"
/blast_score=0.0
/pfam_match="ketoacyl-synt"
/pfam_match="ketoacyl-synt_C"
/pfam_match="Acyl_transf"
/pfam_match="SAM binding"
/pfam_match="adh_zinc"
/pfam_match="pp-binding"
CDS 14275..15408
/blast_score=e-104
/blastp_match="AF285636.WCBT Burkholderia mallei"
/gene="acyl-CoA transferase WcbT"

```

22/39

```
/note="putative"
/pfam_match="PF00155; aminotran_1_2; 1"
CDS 15436..16245
/blast_score=5e-11
/blastp_match="TTDEFFMT.FMT T.thermophilus"
/gene="formyltransferase"
/note="evidence experimental"
/product="methionyl-tRNA formyltransferase"
/pfam_match="PF00551; formyl_transf; 1"
/pfam_match="PF02911; formyl_transf_C; 1"
CDS 16287..17384
/note="putative"
/gene="glycotransferase"
/blastp_match="AF285636.WCBD Burkholderia mallei"
/blast_score=e-99
CDS 17427..18158
/blast_score=2e-82
/blastp_match="AF285636.WZM Burkholderia mallei"
/gene="ABC-2 transporter Wzm"
/note="putative"
/pfam_match="PF01061; ABC2_membrane; 1"
CDS 18248..18847
/blast_score=7e-61
/blastp_match="AF285636.WZT Burkholderia mallei"
/gene="ABC-2 transporter Wzt"
/note="putative "
/pfam_match="PF00005; ABC_tran; 1"
CDS 18952..20346
/note="putative"
/gene="glycosyltransferase"
/blast_score=e-101
/blastp_match="AF285636.WCBE Burkholderia mallei"
/pfam_match="Glycos_transf_1"
CDS 20442..21167
/note="putative"
/gene="unknow"
/blast_score=1e-26
/blastp_match="AE009248.ATU3189 Agrobacterium tumefaciens"
complement(21164..24301)
/note="tranporter domain"
/gene="unknow"
/blast_score=5e-29
/blastp_match="AE009122.ATU1658 Agrobacterium tumefaciens"
/prosite_match="PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1"
CDS complement(24351..27023)
/note="none"
/gene="unknow"
/blast_score=5e-29
/blastp_match="AP003581.ALR0267 Nostoc sp"
complement(27806..29686)
/note="none"
/gene="cell surfarce protein"
/blast_score=2e-34
/blastp_match="AE010748.MA0851 2567 Methanosaarcina
acetivorans"
/product="cell surface protein"
CDS complement(29535..30872)
```

23/39

CDS /note="none"
/gene="cell surface protein"
/blast_score=2e-36
/blastp_match="AE010748.MA0851 Methanosa

CDS complement(30848..32647)
/note="fragment"
/gene="O-antigen"
/blast_score=1e-72
/blastp_match="AF105060.RFBC Riftia pachyptila"
/product="O-antigen biosynthesis protein"
/pfam_match="PF00535; Glycos_transf_2"
complement(32574..35555)
/note="putative"
/gene="glycosyltransferase"
/blast_score=7e-46
/blastp_match="AE013462.MM2213 Methanosa
CDS complement(35533..36598)
/blast_score=7e-37
/blastp_match="AE013462.MM2213 Methanosa
/note="putative"
/gene="glycosyltransferase"
complement(36516..37400)
/blast_score=8e-22
/blastp_match="AP003581.ALR0267 Nostoc sp"
/gene="ALR0267"
/note="putative ATP/GTP-binding protein, esterase fush
9e-14"

gatccccacga tcggcgccag	cgaggatgcgc	agctcgtgcg	cgagcatcgc	gaggaactcg	60
gtcacgtggc	ggccctcgcc	ctccagctcc	tccatgcgt	tgccgtgcga	120
gtcaccttgg	cgaaagcccat	gtgccttcca	ctcgctcgc	gaagcggccgt	180
tcggccaga	agcgcgtgcc	gtccttgcgc	acgcgcgc	ccgtgtcctc	240
tgcaagcagcg	catcgatcg	ctcggcttcc	ggggggccgg	ccgcggcatc	300
tagaagagcg	aaaagtgcgc	tccgatgatc	tcttcggcgc	tgtagccgt	360
gccccttgt	tccagcttgt	cacgtggccc	gccccgttcca	gcatgttagat	420
tgcacggcc	cgatcaggag	gcccgcgc	tcctcgctt	gcccgcgc	480
tcgcggcgct	cgttcaggcc	gcccgtgacc	tttccgaagc	cgaccacgt	540
tcgcgcagcg	cctgtatgtat	gacgttggcc	cagaagcgcc	ggccgtcctt	600
cagcccttgt	cctcgaagcg	gcccgttccg	cgtgcgcgc	gcaagtcctt	660
tcgcgcgcga	tgtccttccct	ggtgttagaaac	gcccggaaat	ggccggccgt	720
gcccgcata	ccttcaactg	ttccggccccc	gcgttccacg	aaacgcacgt	780
tcgagacga	agatcgcgta	gtccgtgacg	gaccgcacca	gcagccggaa	840
tcgcggccca	gcccgcctgg	ggcgcgaag	gcccggatgc	gcaccggagg	900
gatgggcttgc	gcatggctat	tagggatcaa	cggtgcacgc	aatggtcaat	960
gagataattgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	1020
tcgaggccca	gggcctgtac	aatggatggc	tttgcgttgc	tttgcgttgc	1080
tcgcagggtgc	cgccccacgc	gtgtcaact	tcttcgcac	tttgcgttgc	1140
accacccggc	cttgcgttgc	ggcgcacgg	tttgcgttgc	tttgcgttgc	1200
cgtcgggtcg	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	1260
cgcgcttcc	gggcacccgg	tttgcgttgc	tttgcgttgc	tttgcgttgc	1320
ggcttcgtcg	gacgtgtcg	tttgcgttgc	tttgcgttgc	tttgcgttgc	1380
cctcgatcat	cgacggcg	tttgcgttgc	tttgcgttgc	tttgcgttgc	1440
acatggcgag	cctcgaggcc	tttgcgttgc	tttgcgttgc	tttgcgttgc	1500
tgatcgccac	cgacggcg	tttgcgttgc	tttgcgttgc	tttgcgttgc	1560
ocgagacttgc	cgggcgctac	cacgtatgg	tttgcgttgc	tttgcgttgc	1620

24/39

tcctcaccgg	cacgctcgcc	aaggccctgg	gcccggcttc	cgccggctac	acggcggca	1680
agcgcgagg	gttggcctgg	cttcgcaacc	gctcgccccc	ctatctttc	tccaacacgc	1740
tgtatgcccgc	catcgccggc	gctcgctca	aggtgctca	tctctcgag	ggcggcggcg	1800
agctgcgcgc	gaagctcgcg	cgcaacgccc	gccacttccg	cgccgagatg	acgcgcctcg	1860
gtttcacgct	ggccggcgcc	gaccatccga	tcatcccgt	gatgtggc	gaggccccgc	1920
tcgcaagga	gatggcggac	cggctgctga	aggaaggcat	ctacgtggc	ggcttctcg	1980
ttcccgttgt	gccccgaggg	caggcgcgca	tccgcacgca	gatgtcgcc	gccccatgaac	2040
cgaagcacgt	cgatcgcc	atcgccgcct	ttgccaaagt	cgggcgcgat	ttgggagtc	2100
ttgcttggaga	accctttcg	agacgaagcg	cgagcccgc	atctggatgg	tcgagtgc	2160
caagcccg	gtcgccaca	acgacgtgt	gatccgcgt	aagaagaccg	ccatctgcgg	2220
caccgacatg	cacatcttca	actggacga	ctggtcgcag	aagaccatcc	cggtgccgat	2280
gacggtcggc	cacgagtacg	taggcatgg	ggaggccatg	ggccaggagg	tgcgccgcct	2340
gcagggtcggc	cagcgcgtct	ccggcgaggg	ccacatcg	tgcgccatt	gcccgaactg	2400
ccgcgcggg	cggccacc	tgtccgcaa	cacgcaggc	gtggcgtga	accgccccgg	2460
cgcgttcggc	gactacctgg	tgattccgc	ggagaacgcg	ttcccgattc	ccgacgacat	2520
ccccgacgag	atcgccctcg	tcctcgatcc	gttcggcaac	gccccgcaca	ccgcgtctc	2580
gttcgacactg	gtggcgagg	acgtgtgt	caccggcg	ggccgatcg	gcatcatgc	2640
cgcggccatc	gcccgcac	tcggcgcgc	ccacgtcg	atcacggaca	tgaacgacta	2700
ccgcctggcg	ctcgccacga	agatgggtc	cagccgcgcg	gtgaacgtat	cgaaggaaaa	2760
cctgaaggac	gtgatgcgcg	agctcgcat	gtcgagggc	ttcgacgtcg	gcatggagat	2820
gtcgccgtg	ccctcgcc	tccgcagat	gtcgacacc	atgaaccacg	gcccgaagat	2880
cgcctatgc	ggcatcccg	ccagcgaggc	cgcgatcg	tggacccagg	tgatcttcaa	2940
gggcctggc	atcaagggcg	tctacggcg	cgagatgtt	gagacctgg	acaagatgt	3000
cgcctatgc	cagagtggc	tcgacccctc	gccccatgg	acgcaccg	tcgacgtgc	3060
cgcgtac	aaggcgctcg	agacgatgg	ctccggaaa	tccgggaaag	tcgtccgtc	3120
gtgggattag	cgagggccac	ggatcgccg	ccggttcccg	cgccggcgat	cgtcaggcg	3180
gttggcagc	acgccccgg	atttccg	tggctca	gcgacttcgc	cgtcga	3240
ccgatcgca	tcgacgtcg	cgtgagcgt	tcgctgtcg	atgcccagg	cacccgt	3300
tggacgc	cgatggccgg	cagcggcg	gacaccccg	ggctaccgac	ggcgcacgtat	3360
cgcgcctccc	tgccctcg	gccgttcc	ttcccgccg	cgccgcctc	gtcgagttc	3420
gcgcctcg	tgccgtgtca	agtgagc	cgcg	ccacccg	gggcgttg	3480
cccgccggcg	caccccacgg	cgtggcg	ggggctgg	acccgaggc	cctggatgg	3540
acgcggcgc	tcgagacgt	ggcctgtcc	gatccacgc	acagtggtt	ctcccagcat	3600
ttcgaccatg	cgtcgcc	gtgcgtcg	tacctggcg	ccggatggc	cgggtggc	3660
gggcgcgtc	tgatgtcg	atgcggc	ggcatcacgg	cgctcg	cgcgtcg	3720
tatgcggcc	agcagggtgt	ggcatcg	ccggggcg	gctaccgg	gttcccac	3780
atcctcgac	gcccaccc	cggccac	gctgtccgg	acaacctg	gttcctccc	3840
gcccgtca	acgcgtccc	gttcggc	agcagttcg	acgtcatcg	ctcggtg	3900
gcccgtac	atttctgtgg	cggctac	ccctcg	ccgaggcg	gcccgtg	3960
aagcccg	gctgtgtgt	catccatccc	gagctctact	acaccg	ccacggccat	4020
caccccg	agtagac	cgagcc	ttccac	tgaatgc	cgacgagg	4080
cgcacatcg	tgttccgg	ggacgtg	ctccac	cgccgg	cgccgc	4140
cgcgcgg	acttgcgt	gtacaac	ctcaatcg	tcacgg	gcagttcg	4200
gacgagg	gcccgtc	cttcg	tgccgt	cactgc	cgagccgt	4260
gtggag	taca	agccgg	gtcg	cgccgt	ctccgag	4320
tacgtgg	gcatca	caagcc	cgat	cctacagg	gcatgg	4380
gcccattcc	ggtaa	gcccgt	tcagat	ccatagg	ttgttat	4440
cgtcactccc	ggact	ggacgg	gatc	agatcacc	gacc	4500
caggccgtc	tcgccc	cctgt	ccacg	ccacg	ccgcgc	4560
gccgtcac	tgacgg	gagc	ccgc	cgcc	gaaggc	4620
ggcacgg	tcgt	gagg	cgtc	ccatgg	caccgg	4680
accgcac	ccag	ccat	accg	agatcacc	cgtgg	4740
ccgttcgt	ccac	ccac	gggg	ccacgc	cgag	4800
gcccgt	cccg	ccac	gggg	ccacgc	gatgac	4860
gcccgt	cccg	ccac	gggg	ccacgc	cgagg	4920
aaggccgg	agg	ccag	gggg	ccacgc	aaaa	4980
aaggccq	aag	ccat	gggg	ccacgc	gattca	5040

25/39

ttcgacccccc	tgcacccat	gcaggtgcgc	taccaggctg	cgctacgccc	cgaccgaact	5100
tgaaattata	cctgcatcg	ggtcgcagcc	gccatgaatc	gaaacgcccc	gggcttcggg	5160
tagcatccgc	ccatggcgac	gaccctcg	acgggtggtg	caggctacat	ccgcacccac	5220
atccctctcg	cactggccca	ggccggcga	cgagcatct	gcatgacaa	ctactccaac	5280
agctcgccgc	gttccatcg	acgcgtctg	cagatcgac	ccggttgcgt	cgaggccttc	5340
gacgtggaca	tcccgacgc	cgacggcata	cgcaagggtga	tcgcaggcg	cgacgtggac	5400
agcgtgatcc	acctcgccgg	actgaaggcg	gtcggcgaat	cggtcagca	gcccggagcgc	5460
taccacgaca	acaacgtccg	cggcaccgag	agcctgctcg	cggcgtcgc	cgactcgccg	5520
gtgcgcaagt	tcgtcttcag	ctcctcgcc	accgtgtacg	gcctcgccga	gaagatgccc	5580
atcgacgagg	acgcgcccac	ctcgcccgag	agcccgta	gccagaacaa	gctcgacatc	5640
gagcacatgc	tgtcgccct	cgccaagcgc	gaccctcct	ggcgcgttagt	gaacctgcgc	5700
tacttcaacc	cgtcgccgc	acacgagac	gcgcgtatcg	gcgaggaccc	ggccgggggtt	5760
cccaacaacc	tcatgccc	cgtgtccag	gtcgcctcc	ggcgcactgaa	ggagctgagc	5820
gtctacggca	gcgactaccc	cacccggac	ggcacccggca	tgcgagat	ttatccacgtc	5880
tgcgacctgg	ccgaaggcca	cgtggccg	ctggaaagcac	tcgcgcgc	ggccccccggg	5940
accgtcctca	cgtgaaatct	cgggacgggg	ctcggtcata	ccgtgcgtca	attgtctgaa	6000
accttcgagc	gcgtgaacaa	actaagcatt	gccccggcga	tcgtggggcg	gcggccggc	6060
gacgtcgca	tctgctacgc	gaacgcgggc	cgcgcgaagg	ccgtgcgtgg	ttggaccgcg	6120
aagcgcggaa	tcgaggagat	gtgcccgc	gcctggcg	ggcagggaaag	aatgtcaaa	6180
ggctacggtt	gacggattgc	aaatcatggc	cttggcggacc	tttgcgtca	tgccgcata	6240
agggggccgt	tcttgtaa	tccaaagg	gggatgcgtc	caggacgtca	aaaacagcac	6300
ccggcgttgg	aaggaagaag	gcaacccat	tgttgtca	gctccagacg	ggattgcgcg	6360
ccgagtaaga	atcgtcagtc	gagaccact	cgcctccggg	tgcaacgaat	cgtgagccga	6420
agcaacac	cgcacatcg	cgcgtcgt	gcattgcgt	tccgccttc	cgccgtacg	6480
cgcaatccc	tctggccctc	gctcgtcg	gggcgcac	tcgtcaccac	cgtcgatccc	6540
gcfgcgtgg	cgaaggagcc	gtactaccac	ccgcgcac	ccgagccgg	caccagctac	6600
tcgttcgc	ccggcaccat	cgccgcac	agcggcttc	acgcgtcg	ttcgccatc	6660
tccccgcg	aggccgcgca	gatggatcg	cagcgcgc	tgctgcgt	gctctcc	6720
gaagccctcg	aggccgcggc	ggtgaagccc	tcgaccctcg	ggggcagcga	ttcgccgt	6780
ttcgtggcc	tgtccaccgt	cgtactacagc	tacggcctcg	cctacgac	ggcgtccat	6840
gacgcgtcg	cgccaccgg	caacaccgc	agcatgcgg	ccaaccgc	ctcgatattc	6900
tacgac	acggccccc	catggcgt	gatacggcg	gctcgcc	gctgggtgc	6960
ttccaccagg	cctgcacgtc	gatgcac	ggcgagac	gccaggcg	gttggccggc	7020
atcagcctgc	acctgcatcc	gtacgggtt	atcgctt	ccaaggcgtc	gatgcgtc	7080
aagcgcggc	cctgcgcgt	gttcgacg	agcggcgat	gctacgtcg	ctcgaaaggc	7140
gcggccgtgc	ttgtgtgaa	ggacccgt	cgccgcgt	ccgacggca	cccgatc	7200
gcccgtcg	ccgcgagcg	cgtgaaat	gcggccgt	cctccagc	caccgtgccc	7260
agcgtcgat	cccaggctc	attgtgtcg	gatgtgtac	agcgcgcgg	catcgacc	7320
tccgagatcg	actacctcg	ggccacggc	accggcacc	ttgtggcga	tccgatcg	7380
acccgcgccc	tcggcatcg	gtcggccgc	ctgcgc	gccaccagcc	gctgccc	7440
gttccgtga	agagcaac	cggcac	gaggccgt	cgggcatgg	gggcctgt	7500
aaggcgtc	attgcctca	gcaccgc	gtccgc	ccatccac	cgattgc	7560
aacccaaca	tccactcg	cgagtgg	ctgaaggc	tcctcg	gctcacgtc	7620
catccgtaca	agaagctcg	cgtcgg	aactcg	gttccgg	cgccaaacgc	7680
cacgtatcc	tcgagagc	gcccactcg	ccaaaggc	tcgaggccc	ggccggcg	7740
gtggcgtca	tgctgagc	ccacgac	cgccgc	gcgacgtgc	ggcctcg	7800
gcaacttcc	tgccaggc	cgagcaac	cgctgcac	acatcg	caccac	7860
ca	actggc	gctccgc	atgg	gcacc	cgccag	7920
gcccgc	tcgc	cgcc	gtt	cg	cggt	7980
gctgc	cccg	ac	gtt	gtt	gtt	8040
gcaatgg	ccgg	cc	gtt	gtt	gtt	8100
gacgcgt	ccgg	cc	gtt	gtt	gtt	8160
cccg	ccgg	cc	gtt	gtt	gtt	8220
ggcttgac	ccgt	cc	gtt	gtt	gtt	8280
gttgatcg	ccact	cc	atcc	ccca	ccgcgtc	8340
gttgatctacc	accgc	cc	atcc	ccca	ccgcgtc	8400
gttgatctacc	accgc	cc	gagac	cc	ccgcgtc	8460
gtcgcc	ccgg	cc	gagac	cc	ccgcgtc	8460

26/39

gtgaacgtcg	cgtgcataa	cagcacgcgc	aacgtgacgc	tcgcccgcac	gcgcgcccggc	8520
atcgaggcgc	tcgaggccga	gctcacgcag	cgcagggtgt	tccaccgcgc	cctcgaccctc	8580
gactacgcgt	tccacagccc	ggcgatggac	ccgggtgcgcg	acgggtctgt	gagcgcgctg	8640
cgccggcctca	cgccgcgcgc	cacccgggtc	gcgttccatt	cggcggtcac	cggtcgccc	8700
gcccggggca	accagctcg	cgccacccat	tgttgagca	acatccgcga	gcccggtgccc	8760
ttccaggccg	cgatccgcgg	catcgccgag	tcggggtca	atgtcttcat	cgagatcg	8820
ccgcaccgc	tcctcaagaa	ctacatcaac	gacggcctgc	gcccgcctc	gatcgaggc	8880
cgcgcgctgg	tcacgctgca	gcgctcgccg	agcgatcg	cggcgatccg	cgccgcggcc	8940
caggaagtgc	tgatcaccgg	ctgcccgtg	gagaccgcga	agatgtccc	cgcgcaggc	9000
cacttcgtcg	agctgcccgc	ctacccgtgg	cagcgcgagc	gccactggcg	cgcgcccacc	9060
tcgcaggcct	acgacccat	ccagcacggc	aagcagcacc	cgctgtggg	ctatcgcc	9120
cacgagaacg	acttccagtg	ggaaaaccac	atcgacaccg	cgctctatcc	ggcgtacgcg	9180
gaccatgtcg	tcggcggcgc	cggtgttcc	cccgcggccg	gttgcgtca	gatggcgctc	9240
gcccgcctcg	cgatcgcgct	gggcggcgac	gcccacgcga	tcgagacgt	cgagatccgc	9300
agcgcgctgc	tgctcgagga	ctccacgtcg	aagaccgtgc	gttgcgtct	gaaacccgat	9360
ggccgcgttca	caatccgcag	ccgcgcgtcg	ctgagcgagg	atccgtggca	actccacgtg	9420
gtcggcaagc	tggtcgccac	gcccacccag	cttccgcgcg	cgccgcgc	gtcggcgcgc	9480
tcgcgcgaagc	ccgatgtgtat	cgccgcgcag	cactacgaag	gcccgcgaa	gcccggcc	9540
gcctacggcc	cgcccttcca	gtcggtctcc	aagggtgtggc	tcgacgcgc	ggacccgggc	9600
agcgccctacg	cgccgcgtat	cctccccaag	cgcattcg	gcaagctggg	cgtgatgcac	9660
ctgcaccccg	cgtcgttgc	cggtgttcc	cagctgtgg	tcgacactgt	cgccgcgc	9720
gcctcccccgc	acgcgcagat	cgccctcg	ccgatctgg	tcgggcgcac	ggccctctac	9780
ggccctgcgc	ccctcg	cgccgcgc	gtccgcctga	ccgcgcgc	ccgcgcctcg	9840
ctggtcgcgg	acttcgagct	ctacggcccc	aacggcgacg	tggtggccgc	gtcaccggc	9900
gtgcgttcc	cgccgcgtca	gctgaagcag	ccgcaggcgc	cgccgcgtcg	ctacctgtcg	9960
taccgcgc	tcccgccct	aatggccgc	gacacgcgc	gcccgcgc	gctgcgcgc	10020
gcgttgcgtgg	ccgcggcg	cggcgagcgc	tggattcg	tctcgccgg	ctcgcgcaag	10080
cgctactacg	atgagggtcg	gccgcgttcc	gacgtgtcg	gcagctctt	cgccgagcac	10140
gctgctgcgaa	agcgatcg	cgaggcccc	ctcatcgacc	cgccgcgc	cgccgcgtcg	10200
acgcggccgc	agcatggcc	gttcctcaag	cgcgcgatcc	agatgtgg	cgaggaccag	10260
ttgctcgagc	ccgcgaagg	cggtggcg	tggagcccg	tcacccaa	gcccgcgc	10320
caggagagct	ggatcacgt	gctgcgcac	tacccgcac	aggccggcca	gtgatcg	10380
ctcgccgcgc	ccggcg	cctcgccgaa	gtgttcgc	gcaaggaa	gtccttgc	10440
agcctgttgc	cgccacgc	ggcgaacgt	ttcgccgc	tgtgtcg	cgacccgc	10500
tcgcgcac	ccggcg	catcgccgac	acgctggcg	tcgcgc	gcccgc	10560
gcgcacccgc	cgctcg	gctcgagg	acgcgcgc	gtcccg	gacgc	10620
gtccctcgagg	ccgtgg	cgaccgtc	gagtaccac	tcggcg	cacggac	10680
gcactggcg	agtacgaagg	gttcctcg	gccattccgc	acgtcg	ttcg	10740
gacctgaaga	tcccgccat	gggcctga	gcccgc	gatgtcg	cgatgtcg	10800
atcggttccg	atggcctg	cggtcg	gatccgat	cgccgc	gacac	10860
ggcgcgctcg	cggaagacgg	gtgtgttgc	atggtcgc	agcatccgt	gctgtgg	10920
gacctccct	tccgcgc	ccggcg	tggacgtcg	ggccctcg	ctcg	10980
tcgcgcctgc	gttcgc	cgagtggcg	atcg	ccac	ctcg	11040
accgtcg	tggcc	cgccgc	gttc	ccac	ctcg	11100
cgcgacggcc	cggtgc	ggctgc	gagc	ccgc	ctcg	11160
ttcgtgc	ccgc	gctgc	cgca	ccgc	ctcg	11220
cgcgacgc	gca	gttc	cgat	ccgc	ctcg	11280
gcccgttgc	cgat	cgat	cgat	ccgc	ctcg	11340
gcccgttgc	cgat	cgat	cgat	ccgc	ctcg	11400
gcccgttgc	cgat	cgat	cgat	ccgc	ctcg	11460
gcccgttgc	cgat	cgat	cgat	ccgc	ctcg	11520
gcccgttgc	cgat	cgat	cgat	ccgc	ctcg	11580
gcccgttgc	cgat	cgat	cgat	ccgc	ctcg	11640
gcccgttgc	cgat	cgat	cgat	ccgc	ctcg	11700
gcccgttgc	cgat	cgat	cgat	ccgc	ctcg	11760
gcccgttgc	cgat	cgat	cgat	ccgc	ctcg	11820
gcccgttgc	cgat	cgat	cgat	ccgc	ctcg	11880

27/39

gagggtgcgcg	ccgcgggcct	caacttccgc	gacgtgatgt	acgcgatggg	cctgctctcc	11940
gacgaggcgg	tcgagggcgg	cttcgcggc	gcctcgctcg	gcatggagct	ctcgggcgtc	12000
gtcaccggcg	tcggcaagga	cgtgacctcg	gtcgcgcgcg	gcgacgaggt	gctcgcttc	12060
gcccgcctcg	cgttcgcac	gcacgtggc	accaccgcgc	attcgtcgc	gaagaagccc	12120
gcccgttgg	cgttcgagtc	cgccggcacc	atccccgaccg	cgttcttcac	cgtgtactac	12180
tcgttgaagc	acctggccca	gctgcgcgag	ggcgagcgcg	tcctgatcca	ccgcgcgcgc	12240
ggcggcgctgg	gcatcgccgc	gatccaggc	gccaagtggc	tgggcgcgga	gatcttcgccc	12300
accggccgct	ccgacgagaa	gcgcgacttc	gtgcgcctgc	tgggcgcgga	ccacgtgc	12360
gactcgccca	cgctcacgtt	cgccgacgac	gtgctgcgcga	tcaccaacgg	cgagggcgtc	12420
gacgtgggtc	tgaactcgct	ctccggcag	gccattgcgc	gcaacctgcg	cgcgcgtgc	12480
ccgttcgccc	gcttcatcga	gctcggaag	cgcgactact	acgagaacac	gcacatcg	12540
ctgcgcccgt	tccgaacaa	catctcgac	ttcggcgtcg	atgcccacca	gctcatgaag	12600
gagcggcccg	acctcgcgcg	ccagcttcc	aacgagctga	tgcagctt	cgagcagg	12660
gtgctctcg	cgtgcgccta	ccgcgccttc	gcgcgcaccc	aggcgttgg	agccttccgc	12720
tacatgcage	actcgcgc	gatcggaag	gtgggtgtcg	cgttcgcgc	ccgcgtgaag	12780
gcccgcgcgc	cccgccccgc	cgtgcgcgc	gagctggcgc	tgcgcgcgaa	ccgcacgtac	12840
ctcgtaaccc	gcgcgcgtgc	ggcgttcggc	ctgcgcaccc	cgcagtggct	gttggacaag	12900
ggtgcgcgc	acctgggtgc	gttctccaag	agcggcgccg	agtcgtcga	gaccaagg	12960
gcccgtggcc	acctcgaggc	ccaggcgcc	acggtgatcg	cggaggcg	cgacatcacc	13020
gaccgcgcct	ccgtgcaacg	gctgcgcgc	gaggtgcgc	ccgcgcgc	gccgcgcgc	13080
ggcgtatcc	acgcggcg	cgtatccag	gacggcttca	tcaccaacat	gacggccgc	13140
caagatccgc	acgtgtgc	gccaagg	ctcggcgccc	gccacatgg	cgagctca	13200
cgccgcgc	agctcgatt	cttcgtgt	tactcg	cgacgacgt	ttcggcaac	13260
ccggccagg	ccaactat	cgcggcaac	tgcttctcg	aggcaactcg	caaggcg	13320
cgcgtcgagg	gcagccgc	gctgtgcgt	ggctggggcg	cgatcggg	ctgccc	13380
ctcgcgcc	acgagaagg	gaaggaagc	ctgcagacgc	acatggcg	cacccgactc	13440
gagtcggagg	ccgcgtcg	ggtgctgg	cagtcctc	tcgcgcacgc	ctcgggc	13500
ggcgtgc	acttcagct	gcaacatc	cgccgcttcc	tcccacccgc	cgccgcgc	13560
cgttcccg	agctggcg	gcccggc	gacgcgaagg	acgacgacga	ccggcg	13620
gagctgcgc	gcctcgccaa	cgagctgg	gcctccgagc	tctggcgat	ttcaccg	13680
ctgctgcgc	gctcgcc	cgagatc	cgcatcccgc	cgaccgc	cgatcgc	13740
cagcccttgc	aggaaatgg	catggattc	ctgatggcg	tggagctg	gaccgcgt	13800
gaggcgcgc	tcggcgtgaa	cctgcccgt	atgtcgct	ccgagcagcc	gtcgatcg	13860
aagctgg	accgcac	gctcgctg	aaggacccg	acgcggc	cgaggcc	13920
tcgcgc	accagatc	gcccgtc	gcccgc	cgcccg	cgactcg	13980
caggtcgagg	agctcac	ggcggtc	gacgcga	ccacgc	gggacgt	14040
tgaaggcg	caacccgc	ggcctc	aggcagt	ggagcag	atccagcg	14100
tgctcg	gcccgtgc	cggtcg	aggaca	ctcgccgt	gagtcc	14160
tcaacgc	ccgcccc	ccgatccc	tggcc	ctcgagc	tactgc	14220
tcgacatc	cccggtct	cagcagatc	agctgg	ggaagg	gcaagct	14280
gcattcc	tccgtact	cgcg	acggcg	cgccgata	acgcgc	14340
ccggcg	gtacatca	ttctcc	actac	cgcttctc	ggcgcac	14400
ccgtgg	cgccgca	cgccat	acgaa	gaccctcg	ctcgcc	14460
aggctcg	ccggcg	ggcg	cgccg	cgccgc	cgccgagg	14520
tatggcg	acgacgc	ggcctt	aggc	cgacca	ctccaccat	14580
ggccac	tcggcc	cgac	gtgcac	cggtcg	caacag	14640
ctgatgg	tccag	gggc	cgatgg	cgttcc	cgactgg	14700
gccctgg	agctgt	ccgc	ggccat	tcgcgt	gatcg	14760
gagggcg	acagcat	ccgg	ccgc	agcgcgt	tcgcgc	14820
cgccgc	gcgc	gatgg	gacgc	ccttcc	gatggccc	14880
cgcgg	gcatcc	ccactt	atggcc	acgagg	catctggat	14940
ggcac	ccaagac	cgcc	ggcg	tcgcgg	gaaagg	15000
gtcg	tgaagg	ccgg	ttcc	gcgtcg	ggcccc	15060
gtggcg	cggc	cgcg	gac	cgcgtcg	ggcc	15120
gccctgc	aggc	cggt	ccat	gcag	ccgcgc	15180
gccctgc	aggc	cggt	ccat	cgat	ccgcgc	15240
cgcc	aggc	gatgtt	ccagcg	agccgat	ctacccgg	15300

28/39

gtgcaggaga	actcggcgcg	cctgcgcttc	ttcgtgagcg	ccacgcacac	cgaggagcag	15360
ctgcgttca	cggtccgcga	gctcgccgac	gcctggcgca	agctctgagt	ggcgccccc	15420
cggctgcgcg	tcgtcgctg	caccacggc	ggcctgaacg	gcgcgtgtt	gctctcgcg	15480
ctgctggccg	cccccacgt	cgaaagtctc	gcgcgtgtt	tctccagccg	cgcgccggt	15540
gccccatgagt	cgtatggccg	tgccgcgctc	ggatacgtgc	gcgcgagccg	cgtcgctat	15600
gcgcgtgtacc	tctggtgcgc	cacggcgctc	gccgacctgc	tgctgcgcgg	cacgtccgag	15660
ggcccggtcg	cgcgcatcg	cctcgccgc	ggcatcccgc	tcctcgccac	gccgcgcgtg	15720
aacgatgcca	ccgcgcgcgc	ttcatcgcg	ggggcgccgc	cggacctgtat	cgtctccgcg	15780
ttcttcaacc	agcacatcg	cgccgacgtg	gctgcgctcg	cgcgctgtgc	cgcggtcaac	15840
atccaccctgt	cgcgcgtgc	gcaactccgc	ggcgtggatc	cggtgtcctt	cgcgccctg	15900
cgcgggtccg	agcgccacgg	cgtgagctg	catcgcatcg	aaccggctt	cgataccggc	15960
gcgcgtctcg	cccaggaaac	cgacgtcgag	gcgcgcgcga	gcgtctttgc	cgccacccgcg	16020
gcgcgtctatg	accgcggggc	ggcgctcctc	gcgcgcgcgt	ccgcggcgct	gcgcgcgcac	16080
ccgcgcggaa	ccccccagcc	ggcgggggc	tcctacgact	cctggccac	ccgcgcggccag	16140
gtcgccgcct	tccggcgcgc	cggcggcgc	ctgctccgc	cccgcgacct	gtgggcgcctc	16200
gcgcgcgggg	ggccggccgc	tttcgtataa	gaatcagcgc	ggtagcggcg	actacgcgcg	16260
cgcttttccc	cgaccccccgg	aggccatga	aacactggct	gaagcaacac	gcaatcttcg	16320
cgctcaccgt	cctgctgcgc	accgtggccg	cgatcctcta	tttcgcctcg	atcgccctcg	16380
acgtctacat	ctccgaatca	cgcttcgtgg	tgaggagcccc	ccagcgccag	gtgcagacccg	16440
gcctgggtgg	cgcctgctc	tcgggcacccg	gcttctcg	ctccaggac	gacacacttact	16500
cggtgcacga	tttcatcacc	tcgcgcacgc	cgctgggcga	gctggacaag	aagctcgccg	16560
tccgcaagct	ctacacggcc	gccaacatcg	acttcataa	ccgcgtcccg	gggctcgact	16620
gggacgcacag	cttcgaggcg	ttccaccgct	actaccagaa	gaaggtcacg	atcgacttcg	16680
acaccgcgtc	ctcgatcact	gtgctgcgcg	tgcgcgcctt	cgagaaggcc	gactcgccgc	16740
gcatcaacga	cctgtgtctg	cagatggcg	agcgcctgtt	gaacgagctg	aacgagcgca	16800
gcgcgcaggaa	cctgtatccgc	ttcgccgagg	ccggaggcttc	gctccggag	gacaaggta	16860
aggatgcgcg	gctggcgctc	tcgccttcc	gcagcaacca	gtcgggttcc	gagcccgacc	16920
gcgcaggcgtc	gatccagctg	cagggcgtgg	ccaaagcttca	ggaggagctc	atcgccaccg	16980
agggggcagct	ggcacacgtg	cgcaagctt	cgccgcacaa	ccgcagatc	ggcgccgtcg	17040
agaacaaga	ggcggcgctg	cgcgtggcg	tggcgcgcga	gtccccaag	gtgaccggcg	17100
gcagcggcgtc	gttcagcgcg	cgcgcggcc	cgttcgagcg	cctcaccctc	gagaaggct	17160
tcgcgcaccg	ccagtcgggc	gttgcctca	ccgcgttgg	gaccgcgc	agcgaggcgc	17220
agaggaagca	gcttacacc	gagcgcac	tgcaagccaa	cctgcggac	aatcggtcg	17280
agccgcgcgc	catccgtcg	atcttaccc	tgttcgtgt	gggcctgg	gcctggggcg	17340
tggtgaggct	gttggggcg	agcgtgcgc	agcacgtcg	ctagggcg	gagcagaacc	17400
cttccttcct	gcgcgccttc	gggggtgc	ggcgcgtgt	ctacgcgt	ctgatgcgcg	17460
aggtcatcac	gcgttccgg	cgcgacgacc	tcggcgtgt	gtggctgg	gtcgagccga	17520
tgtatccac	gttcggcg	acggccctgt	ggaccgcggc	cgcatgaac	cacggctcg	17580
cgctgcgcgt	cgtggcccttc	cggttgcaccc	gatactccgc	ggtgtggta	tggcgaact	17640
gcgcctcg	cgcctcgatg	gccatcgagg	ccaaacaagg	cctgccttc	caccgcac	17700
tgcgctgtat	cgacgtgtt	gtacgcgc	tcatcctcg	gatcgccggc	gccacggcg	17760
cgttacccgt	gttgggcattc	ttttcatct	ggatcggt	gatgcgcgt	ccgatcgaca	17820
tgctcaagg	cgcgttcggc	tggttcatgc	tgtcggtt	cgccgcgc	ctggcgctcg	17880
ccatcgccgc	cggcacccgc	tactcgctgg	tgtcgagcg	gctgtggcc	ccgacggcct	17940
acctgctgtt	cccgctctcg	ggcgcgcct	tcatggtca	gtggctgg	gagaaattcc	18000
agcagttcg	gtctttcg	ccgatggtgc	acggcaccga	gctgtgcgc	gacggctact	18060
tcggcaacgc	ggtgcgcacg	caactacgac	tgggtacat	ggcaacggc	tcgctggcc	18120
tgtcgctgt	cggcacatgc	ctggcgc	tggcggccaa	gcccgtgg	gccgcgtgt	18180
ccgggtcgag	aacgtgacca	gatgtacta	cgccgcctcg	cccagaacgt	gctgacgcgt	18240
cagttcg	ctccggccgc	ggcgcacat	cggcacatc	ggcaagaacg	gcccggccaa	18300
gtccacgcgt	atccgcctg	tcagggcgc	cgagagcccc	accaccggc	gcacgtcg	18360
cgagatgagc	gtgtcctggc	cgctcgctt	cgccggcg	ttccagaccc	acccacccgg	18420
gctggacaac	ctgaagttcg	tctgcgc	ctacggcgt	gactaccgc	acaaggctcg	18480
ttcgtcg	gacttcacc	agctggcg	gtatccgc	gagccgg	tccactact	18540
gcacggcatg	atcacgcgc	tcgcctcgc	gtgtcgat	gcgggtgg	tcgactgtt	18600
cctgatcgac	gaggcgatgg	tggtggcg	cgccgccttc	cacgacgc	gccacgtcg	18660
gcttccac	aagcgcaagg	accgcgc	catcctcg	acccacgc	ccaagggtat	18720

29/39

caagctctat	tgcgagagcg	cctgcgtgct	gcacgagggg	cggctgctgc	cttccccac	18780
cgtggacgcc	gcctacgagt	tctacatgaa	cgaggtcatg	caggacctcg	ccccggaggt	18840
tgcctgaacg	cccccgccg	gccgctgcgc	gtgcctctcg	agctgcggcc	ggcgttcgac	18900
ggccacgccc	gatccccca	ggaggcacgc	ctgctgttcc	gcccctgcg	catgatcgag	18960
ggcatcgacg	tggagggcct	gctgcagcac	agcggccacg	tgctcgccaa	gggcctgcgg	19020
ccgcgcggcg	gcggcgacct	cgcacccggac	cgccagctga	accgcctgtc	gcccgtcgt	19080
gtctccacca	agcaggagct	caccaacgcg	cacgtggcg	cggcgtctgat	ggccattcgc	19140
cagctgtctgg	gcgcgcgcga	gaccctctac	cgcttcgacg	cggcgcactt	ccgcgacttc	19200
gtctggcagg	cgctcttcgc	gcccacgctg	cacgcccgg	acttcgactc	gttcacccgc	19260
gcccgccttc	gcgtcgccg	cgtgccgtgg	acgggcatgc	accgcgtcgc	gctggtcacg	19320
cgcaagctcg	gctacacgct	gcatccgcgc	atcgacactt	ccgatttcga	cggtctgatc	19380
gcccagacgc	cctacccggc	gccccgtcg	gacggcacgc	gcccgtgtt	gcccgtaccac	19440
gacgcgatcc	cgctcctcat	gcccacacc	atctccgaca	tgtccatcca	ccaggcctcg	19500
caactaccgg	cgctcgccg	caacgtggcc	tccggcgcgc	acttcgcgt	cgctcccgaa	19560
gccacgcgca	aggacctgt	ctcggtctt	cccgaggtcg	aggcgcgtc	cagcaccatc	19620
cacaacatgg	tctcgccacca	ctacttccgc	gagaccgcgc	acgcggggcg	catcgaggag	19680
atcctgcgaa	cgcgcgccgag	cgagcgcatac	aagggcgcgc	ccaacggagc	cgccaccgc	19740
gtcgacacgt	cgcccttcctt	cctcgccgc	gcccgcacac	ccagcggaccc	cggttacactg	19800
ctcgccgtgg	cgacgatcga	gccgcgc	aaccacgcgg	gactgtatcg	cgctgtggag	19860
cacccgtgcga	cctcgccgtt	ccccggctg	cgctgtgtgg	tggcgtccat	gcccggctgg	19920
ggctacgagg	cgctggtagc	caagttcaag	ccgtggctcg	cgcggggcca	gtcttcgt	19980
ctcgaggacg	tgcccgcccc	cgagctgcgc	ctgtctact	cgcacgcgc	tgccacgggt	20040
tgcccggact	tccggcgaggg	ttcgatttc	tccggcg	aggcgtatcg	ctgcggcagc	20100
ccgggtatcg	cctcgccagat	cgccgcgcac	cgcggatgtt	accgcgcacgc	cgccgagatc	20160
tgcagccctt	atccgtggc	cgacccgc	gaggcgatec	ggcgcgtcat	cgaccggcg	20220
gcccacgggg	tgcgcccaggc	gctggtcacg	cgcggcaccc	agggtatcg	gcccgtacacg	20280
cccgaaggca	tcctgcccga	gtggcg	tacctgtcg	gcacgggtcg	cgcggaggcc	20340
ccgtgagcga	tccgaaggcc	accgcggagg	cgtggggagga	gtggggccag	cgcgaccct	20400
acttcggcgt	gatcacaac	ccgcgttcc	gcccggggca	gatggacgag	gaagcgaagc	20460
gccaatttct	cgctccggc	cgcgtccatg	ccgactacgt	gatgcgtatcg	gtgcacgcgc	20520
acatcgccgc	cgacttcgt	ccgcgcacca	tcctcgattt	cggctcg	gtggcccgcc	20580
tcgtcatccc	cttcgcccgc	caggccgagc	aggtgaccgg	cggcgcacgtc	tccccctcg	20640
tgctggccga	ggctgcgcgc	aactgcgcgc	agcagggcg	gccaatcg	cgcctcg	20700
tctcgacga	ctcgctcacc	gggctggcc	gccccttcga	cctcg	tcgttcatcg	20760
tgttccagca	catcgatccg	gcgcgcgggc	gcaatctt	ccgcccctg	ctgggcaccc	20820
tggcgagcgg	cggcgtgggc	cggttgcact	tcgtctatgc	gaagcgtatc	tacgcggcga	20880
cgtacggcgt	ggcgcaccc	cccgcgcgc	cgcgc	gcccgcgc	cagcctccac	20940
ccagccgggc	ggagatcaag	gccgcgc	aggcgcgggc	cgcgtggcc	acgcaggcac	21000
gggcgcggga	acccgcaccc	aatccgcacc	ccgacatcg	gatgaatttcc	tacccggcgg	21060
cggagatgt	gttcctcg	caggaagccg	cggtc	ccgcacgtc	gatgttaccc	21120
accacggcgg	cgagctgggt	ctttccctt	tcttcgc	gccc	tccttcgagg	21180
cgaaaggagag	gcccacgcgc	gtatcg	ccgcgt	ggggcgacc	gcccgcgc	21240
ccgcgcacttc	ggcagccacg	cgctcg	ccgcgc	ccgcgtacagg	cgacgcgc	21300
agtagtgcac	ccagtaggag	tccggacagg	ccgacgc	tccgc	cgctcgac	21360
cgggccgcac	gcgc	tcacttgg	cgagg	ccgcacgtc	cggtgttagt	21420
gggtccagaa	cgaaggcgcc	cgccatcg	gcgcgcgc	gacctcg	ttgcgtc	21480
ccccggagag	ccgggacgc	gcccaca	gttcgc	cgccgcgt	gcccgcacgc	21540
tccgcaacgg	cgccaccc	gggtggcg	cctcg	accctcg	gacgagaagg	21600
ccgagcgcgc	gaactgcac	agcgat	cgagc	atgtt	cagagccgc	21660
cgggatccac	cgccgcgc	ggcgg	cacgc	cgcc	tcacc	21720
gcccgtgcag	ccgcaggcg	tggaa	ccactccgc	cgacatcg	gcccgc	21780
gctccagcgc	accctcg	gcccac	cgcc	gatgtgt	aggatggccc	21840
aggtcgcctg	cgccgcgggg	agg	ttcg	cgac	aggaccgc	21900
gcccgcggc	ggcgtcg	act	cggtt	gggtcg	aggaccgc	21960

30/39

cgagctcggt	ggccgcgcgg	aaggcggccg	ccgcgtcgcg	ccccaccgcc	gcgagggtcg	22200
cgcgacggc	gtagaggcgc	gcccgggtgt	aatacgccca	gaacgggtcg	gggtcgcgaa	22260
tcaatgtatc	gaacgagcgc	tcgacccctcg	cgccacgtt	gcccgggtcc	acctgggtcg	22320
cgccccacag	ggtcggcgc	tcgacccctcg	cgccacgtt	gcccgggtcc	acctgggtcg	22380
tcaatgtatc	gtcgccgaa	cgccgctcg	gcaacgcgag	ctccaggggcg	ttcagccgt	22440
cgaggctcg	cttgtactcg	cgataggcat	cgccgtcg	gtcgatattcc	tgccggcac	22500
tgtcgatccat	caggcggggc	gtgagccgt	gtcgccgt	aacgggtcg	tatgcgtagg	22560
cccgcgatc	gtcgtagggc	agcgtggcc	cgccgagcga	cccgaacgcg	gcgagggtcg	22620
gaaatcccg	cttggccagc	acgccccggc	acgccccggc	ggcttccatc	tcgccccggc	22680
tccagtcgc	caggcggcgc	acgccccggc	acgccccggc	cgccgtcg	tagagctcg	22740
ccagggtccgc	gtccggggcg	coctgcgcct	cgaaagatcgc	cttccactcc	gcgagtcct	22800
cggtggccgc	ggagaagatc	ggcacgtgt	cgccgaagat	cgccagcacg	tagggccg	22860
cgccgaccc	gagacgctcg	accagcggc	gcagcgcgcg	gtcggttgc	gccccatcacc	22920
cgcagtactt	gtcgatcg	ccctggcga	cgccgtcg	gccccatcacc	gagacgccc	22980
agcgcggc	ggcgcccttg	tcgccccagg	agggtccgt	gccccatcacc	acggccgaaat	23040
acacgaagcg	cgccgcgcgaa	tcgtcgatct	ccgcgtatcac	ggcatccatc	tcgtcg	23100
ccgcgtatcg	ctcgaacttg	cgccggccat	ggcgaacgc	tcgtcg	tgaacgcgc	23160
cgaaggccag	cttcttgcac	acgcggtc	ggtttagaa	ccagcg	ttggatgg	23220
tggccacgt	ggcgtagccg	tggcgccga	gcatgc	cccgatatac	acccggtc	23280
gcaactgccc	catgttagggc	agcgcgcgc	aggcttcc	gccccatc	atgcgggt	23340
gatctcgaa	ctcggtgtcg	cagggttgc	cgccgatcac	cttcacgc	aggccggccg	23400
tgtcgccgc	cgccgcgtc	cgccgcgag	acggcagcg	gtcg	cgccgc	23460
cgaacggccg	caggctggcg	agcactcca	tcatcaggaa	gatgtgtcg	ggcgctcc	23520
cggtggccgc	ggcgctcgcc	gaggctcc	ggcgctcc	cagccgtcc	acccctcg	23580
gcagccgcg	ggcgccgacc	ggcgccgca	tgctcc	gaccacc	gcgtgc	23640
cgaagcccg	gcgcgcacg	ttcacatttgc	gacgaaaggc	gacgcgcg	ggggcg	23700
actccacccg	ccagcggcgc	aagcgtccg	cgacgaaaggc	tcgtgtatc	tcgtgtatc	23760
cgagcagcca	ctgccccgca	tcgtcgatcg	cgctcg	gtcgatggc	tcgtgtatc	23820
cgaacgcggc	ccacgcggc	gcccggcagc	tcgccccggc	ggcaacgcg	ccacccgca	23880
cgccggggcc	atgcggatc	gcccggcga	acaccagc	tcgcgc	tcaggaaat	23940
cgccggggcc	gaccacgcg	ccgaaagacg	cgacggggc	tcgttcc	tcgttagagc	24000
cgccggccag	cgccacggc	cccacccgt	cgacggccc	gaccgcg	aagacgc	24060
catggaaagat	ggcgagcgg	aggccggc	aggccaggc	ccgcgc	gcccgggt	24120
ccacgtata	gcccaccacg	aggccggc	cgaggccggc	ctgcgc	tcagggt	24180
gcacgcac	cagcgtccg	cccacccgt	cgccgggaa	acccgc	ccagcg	24240
cgtcgaggat	catggggatc	ccagggacgt	ttccggcca	gtctataac	ccccggcc	24300
tgaaaaacgg	cgaggccgg	gggcctcgcc	gtggagggg	ccgcgc	tcagggttgc	24360
gggggtgcga	ccgaccacgc	ggccggggcc	atcagggttgc	cgctccgat	gtggcg	24420
ccgttcatca	ggcgcacgt	gacgctccg	tcgtgttgc	ggaagaccc	gtcttgcgg	24480
ctgtcgccgt	tcagggtcag	cagctcg	acgttccaa	ccgtccc	ccgcgc	24540
tcgccccgtgg	tgatgtatgg	aacaccatcc	atcaggccg	tgtgcacgc	ccgtcggt	24600
tggccgaaga	ccaggtcg	cttcgcgtc	cggttcagg	cgcccacc	gttcacc	24660
cacccgtgc	cgccggccag	gagctggccg	ctggcgccg	agggtgt	gttcatcagg	24720
aacagggtcg	cgccggccgtc	ggtgtggccg	aagatcatgt	ccgcgcgt	gtcgccgt	24780
aggtcgc	ggtggcttac	cgtccacgc	ctggcgccg	agaggaagcc	cgctccggcc	24840
gtcaccgtgg	tgccgtccat	gtatgtatgt	tagccgc	cgtcgtgt	catgaagacc	24900
aggtcgc	tgccgtccat	gtatgtatgt	cccgcc	tgacgttcc	gcccgtcccc	24960
gccccgaaga	gttggccgt	gcccgtccat	gtatgtatgt	ccatcaacca	gtatgtcg	25020
cgccgcgtgg	gttggccgt	gatgtatgt	cccgcc	cccgcc	gtccgcgt	25080
tgctc	tcaggccg	cccgccggc	agagcttcc	tgccggccgt	caccgtatgc	25140
ccgttcatga	tgtacgcgt	ggcgccggc	tcgggttgc	ggaacaggat	gtcgccgc	25200
aggtcgc	tgagggtcg	cgctccgtc	accgtcc	ccgcgc	gcccgtcagg	25260
ttggccgac	cggtgtggc	gttgcgttc	atcgcc	ccgcgt	ccgtcg	25320
ttctgc	cgagggtcg	cgtccgtc	ccgtc	ccagacgc	ccgttgggt	25380
ggccgcgtga	cgccaccgg	aacccgg	ccgcgttcc	cgatcc	cggtcg	25440
cagttggcgc	catccgggc	cgccgggtt	cggtcg	cggtcg	agttgcaagg	25500
aagttcgaga	aacgcac	cagcgtgtc	gcccggcgg	tcaccggaa	cgacacc	25560

31/39

ttccagcccg	aggccgtgg	gcccgcgtcg	ctgaacacca	ccgtggcg	tc	gatcagg	aa	25620
tcgaactgc	cggcg	ttgggg	gaag	ctcgac	acgc	gatagg	cgaac	25680
agcagcgtgc	ccgc	gaac	gagg	tcgc	gtcg	gggt	ggcc	25740
gtcaccacct	gcgc	gaac	aagg	tcgc	gcgc	gggt	ggcc	25800
gtccatgcgt	tgg	gcgc	cgagg	tcgt	aagg	ccgg	gagc	25860
cacggcgaga	ggat	gtat	gg	gtcg	gg	ccgg	gacc	25920
cccgtcgcat	tgg	agag	gg	cggt	tt	cc	acc	25980
gccaccgg	ccgt	at	gg	tc	tc	cc	gt	26040
gaggtcgcgg	tgt	at	gg	ccc	gg	gt	gg	26100
gagatggcac	ccgc	cg	gg	gtat	ccgt	gg	gt	26160
tcccgcccg	cga	agg	gt	cg	tc	ac	cg	26220
accgtggact	tgag	gcgc	gg	gt	cc	ac	gg	26280
gcctgcgcgg	tgc	gcac	gg	tc	tc	cc	gt	26340
ttggagaagc	ggta	gac	tt	gtcg	tc	at	cg	26400
ttcgccgcgt	cgtt	cg	tc	gtcg	gt	cg	cg	26460
gagccgttgg	gc	agg	gg	tc	gg	tt	gt	26520
tacgagtacg	cga	ac	gg	tc	gg	cc	gt	26580
gagggggcgt	cgtt	gg	gg	ccat	gg	cc	gg	26640
caaccgaaca	ggc	ag	cc	at	gg	cc	gg	26700
ccgatccagg	ctgc	gc	cc	gg	cc	gg	cc	26760
tccgcgcgt	att	gc	gt	at	gg	cc	gg	26820
gctgcgcgg	gggt	gtat	gg	at	gc	at	gg	26880
acctgcgtcg	catt	cac	ca	gc	cac	gt	gt	26940
ttggcgcgcg	tgac	gagg	aa	tc	gg	cc	gg	27000
aagccgcgcg	agt	ac	ac	tc	at	ac	gtt	27060
gcgagccgac	ggt	att	ca	cg	gg	gg	ct	27120
cgacgcgcga	tttc	gag	cc	tc	gt	gg	cc	27180
gcatgcttgc	ctgc	tgc	gg	tc	at	gt	at	27240
ggcgtactc	acc	ct	gc	tc	gg	cc	gt	27300
cgatgcgcg	gtg	tt	at	cg	tc	tt	cg	27360
cgccgtggc	ctgg	gag	ca	tc	aa	gg	ac	27420
gctgac	tc	gg	cc	tc	gg	gg	cc	27480
gcccggcaac	ggc	g	ct	gg	cc	gg	gg	27540
ggggAACCGC	ggc	ga	gg	ac	gg	aa	gg	27600
ctgtgatccg	tt	gc	gt	at	gg	cc	gg	27660
cgctatgtcg	cgg	ag	at	tc	gg	cc	gg	27720
gggatgtcc	aaa	agg	act	cg	gg	tt	gg	27780
gaactcc	tgg	tct	cc	gc	act	ta	gg	27840
ggccgatcaa	gtcc	cg	ct	cc	at	cc	cc	27900
gcccattgtt	gtgg	tgg	aa	acc	at	cc	cc	27960
gcgtaccgt	cc	ag	cc	cc	tc	gg	cc	28020
cgttcatcag	gc	gc	ac	gt	cc	gg	cc	28080
cgtcgcgtt	gag	at	ct	cc	ac	cc	cc	28140
cg	cg	ct	gg	cc	at	cc	cc	28200
ggcggaa	ag	ac	gg	tc	gt	gg	cc	28260
agccgtgc	gg	cg	cg	cc	gg	gt	cc	28320
atgcgtagcc	gc	gc	cc	at	gg	gg	cc	28380
tgtcgccgt	gc	cg	ac	cc	at	cc	cc	28440
tgatcgccgt	gc	catt	tc	at	cc	gg	cc	28500
gtcggcc	cc	gt	cc	tc	gg	gt	cc	28560
ggggAACAGC	tc	tt	gc	cc	ac	gg	cc	28620
cccg	tc	gt	gg	at	gg	cc	cc	28680
tgatc	ca	cg	cc	gg	at	gg	cc	28740
cg	tt	cc	ac	cc	at	gg	cc	28800
cg	tc	gg	tc	gg	cc	gg	cc	28860
cg	gag	aa	cc	gt	gg	at	gg	28920
acgtgcggcc	gg	cc	ac	at	gg	cc	cc	28980

32/39

cgacgttgtt	gccgcccgtc	agcgtcgaga	acacgatgtc	gccgttgcgc	acgcgagtg	29040
ggaacgcgtc	ggcatcgccg	ccggggccct	gcaccgggt	cagggtcggg	aagccggcga	29100
cctgcgtgt	gcccacgacg	agcgcacgc	cattcggcgc	cagggcgacg	tccagggat	29160
tgtcgtaggc	cgagccgccc	aagatcccc	acgattcgag	cgcgcgggt	gccgagaagc	29220
gcgtcacaa	cgcacatcgacg	gtgccggcga	aattgcac	cgttccggc	tgcgggaagc	29280
tgggcaggcc	ggtatcgccg	acgacgtcg	cctggcccgt	gaccgggtt	accgcgtatcg	29340
cggtcacgtt	gtcgccgctg	ccgcccggag	acgtggaaata	cgcgaccggg	ccgcccgtctg	29400
caccgatctt	ggtgacgaaag	ccgcccaccgg	gttgggtctg	gtacgcattt	acgttgcgga	29460
acgtcgagga	gcccgtggag	cccgcgacgt	acgtgttgc	cgcattgtcc	gttgcgatac	29520
cgcggccat	gtcatcaccg	ctgcccgg	gatacgtgg	atagacgaaag	ccgcccgttga	29580
acgagagctg	gtgacgaaacg	ctgcgcgtt	gaggccgggt	cgcacaaca	cggtgcagc	29640
gcccgtccgc	cggtcgggaa	gccgcgcgc	gccgaaccccg	tgacatgc	gaagccgttg	29700
ttgtccacgg	cgcacccat	gccctgtcg	tgcggggagc	cgcggcag	ccgcccggaaag	29760
atgacgtgc	cagtggatt	cagcgcgt	acaaatgc	cgcgcggc	ggcatgggt	29820
gtgcaatcg	tgacggcgaa	gcaatcg	aacgcgtt	tgcgggtt	accgggtatc	29880
gcccgttgc	cgttgcggcc	ggaggcccg	acgcgcac	cgcggagata	cggttcggcc	29940
ccgagcgtgg	tgcgtacac	cacctgtc	gttgggtgggt	tgaacttac	caccgtggct	30000
tctgtacgtgc	ctgcggatata	gttgcggagc	ccgacgtaca	cgttgcctt	cgcgtccacc	30060
ttgaccggcg	tgttccactg	ctcgtcgcc	gcccgcggcc	gttacgtt	gaacgcacatc	30120
accgggtcg	tcacccacgt	gcccgtctg	tgcgtatcg	cgcggcctcg	30180	
gcccctgt	cgcacccaa	gagctcgaag	cgcgcggcc	cgcgcacgc	cgcacgcgc	30240
acctgtcg	aggctaccgg	cgcgtgtc	gtaaaattc	cctggcc	gacgcgtgc	30300
agggtgcct	cgcacatcgat	ccacgcgtt	tgcgggac	agaggtcg	gcccgttgc	30360
cgcggatcg	cgcgcggggc	gaccacaa	tgcgtatcg	gcgtgcctc	cttgcgtac	30420
acggtgagg	ccacgcgc	atacagg	tgcgtatcg	cgcggcc	gtcggcgcac	30480
ttttcgcg	ggccggcg	cgtgggtcc	ctgtatagt	ggctgagg	ggccgggggc	30540
tcctcg	cgtgaccgg	atcgaa	cccgca	gcacgc	cagcgaggcg	30600
ctggcagctt	tcctggagcg	tggctcg	gcaatgc	cgcgcgtc	ggagacgcgt	30660
ccggcctg	cgcgcgagac	gtagagcg	tccggcc	actggcc	gttgcgtct	30720
aacgtgatgg	gcccgtt	gagggcat	gggacgac	catggac	ggcgggagtg	30780
gcgagggcg	ccgcgac	cacggc	gggcggagc	ccggcgg	gccgctggag	30840
ctgcgggtca	tggggcagt	acctttct	tgttatgg	gttgcgtt	tttgcgggat	30900
ccacagctcg	gatccagct	gatcgtact	cagcgac	gtgggtttt	gtgacccagt	30960
cccgcgtcg	ggcccccgg	acaacgg	catcg	cttgc	gcacgcac	31020
cctcgatcca	gtcgccgt	cgccc	ccgcgggg	atcg	ggctcgagat	31080
ccgaggccac	cacccggat	ccgagg	cgtact	cagg	ttgcgtt	31140
cgcgatggaa	cgggttgc	cgcag	ccaccgc	gtcg	agcgcac	31200
gcttctccg	gtactgc	atggg	tgcgt	ttcc	ggcgg	31260
gctcggcg	gcacagg	aggaa	gcgt	gg	aacggc	31320
ccggctcg	caacttc	tcttcc	gtgtt	cc	ccacgc	31380
gacgcgc	acccgc	cgttgc	ggccgg	acg	tcgat	31440
tcgggatcac	gcccgtt	ttcg	ggcc	ctcc	cgag	31500
acaccacg	gcggcgc	aacgc	cgcg	gttgc	ggcgg	31560
agatcg	cgatc	ttgc	gac	gcgt	gcgc	31620
ggaccttc	cgcgcgt	tggc	gcac	ggat	ccac	31680
ggaagaagtt	gtgcgc	acgg	cg	gttgc	ggat	31740
gcgcgc	cgtggcgt	tcgc	gttgc	gttgc	ggat	31800
gcgcgc	cgctggc	tcac	gac	gcgc	gcgc	31860
cgca	cgatcg	gcgc	ccgc	cgat	ccac	31920
agccctcg	gccatgc	aacgc	ccgc	gcgtt	ccac	31980
gccagcg	gcccattc	gcctt	ccgc	gcgt	ccac	32040
cggcatcg	gcccgc	gcgc	ccgc	gcgt	ccac	32100
ggatgtcg	gcccgtcg	cgca	ccgc	gcgt	ccac	32160
cgcg	ctcg	gcgc	ccgc	gcgt	ccac	32220
gacggcc	cgcgcgtc	acctcg	ccgc	gcgt	ccac	32280
cgccctcg	gcccggcg	gcgc	ccgc	gcgt	ccac	32340
agccgg	gcccggaga	ccgc	ccgc	gcgt	ccac	32400

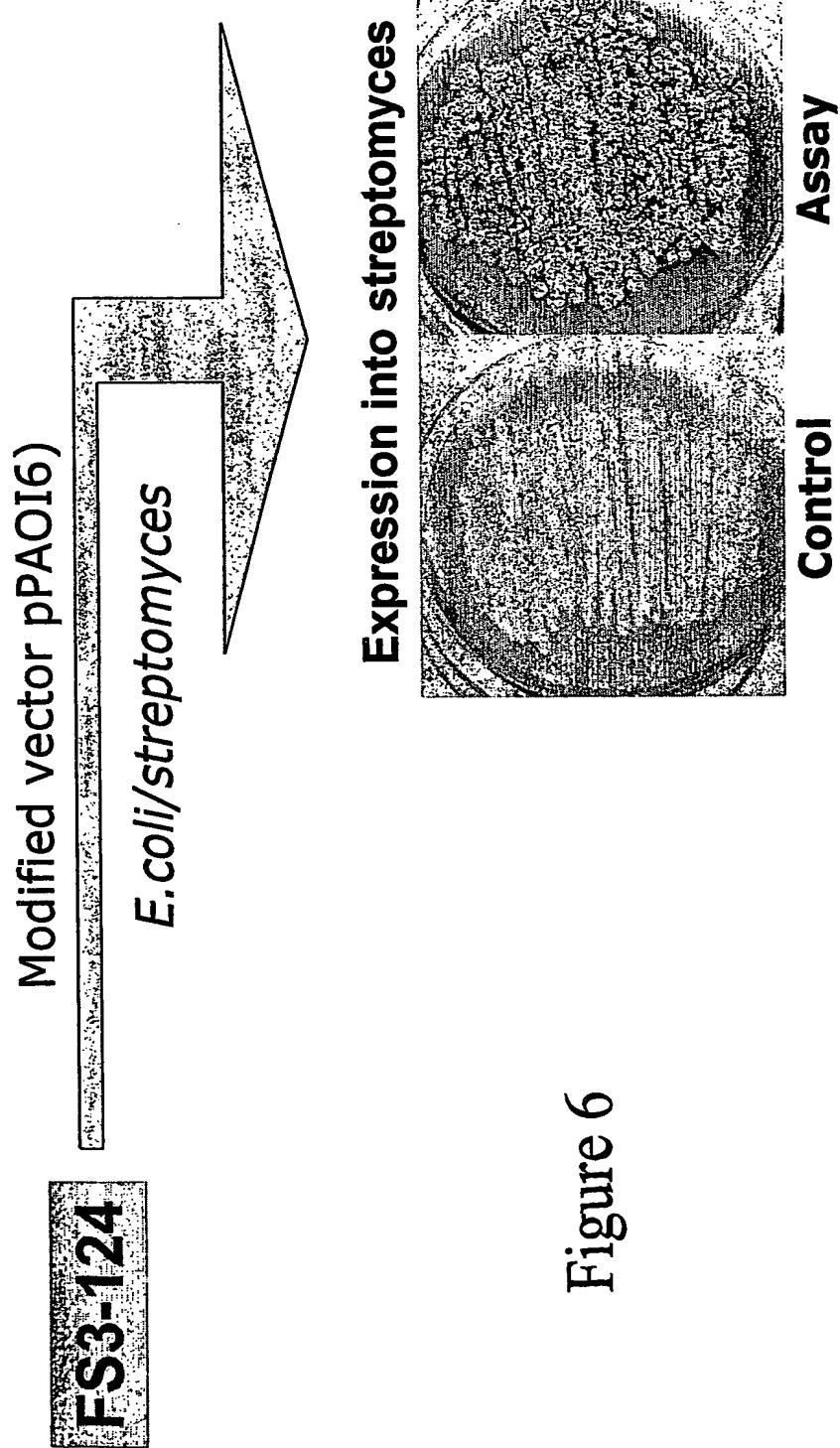
33/39

ggcctgcgat	gccctgctcg	aggcgctcg	gccagccggc	gcggaaattcg	ctgcaacgcg	32460
catcgacat	cgcgatccac	agcgtcgccc	cgttgcggat	ggcgccgcg	aacgcggagc	32520
cggttcgccc	gcccacccgg	accaggtgcg	tgcgacgcgc	gcgcggcc	agctcaccgg	32580
gccagcggtt	cggcgccctt	cgcgtcgaga	tcgacgtaca	cgtggccgat	acgttcgccg	32640
cagcccttca	cgagcgccctc	gatgcacgccc	accagccgct	ccggccgtg	cgtcgctca	32700
atgaaggcgc	aaatcggtcc	ggcccccgc	gcgcggcc	acgatgcggg	cgcateggac	32760
gcagactggc	ggaacgcggc	ctgcgtgc	gtgcggccgg	agtgcctcg	gaccacgcgc	32820
tggtggcgct	cgcgcgcggc	gtcgcttcc	tgcgccggcga	tgttcaggtc	gaggcgatgc	32880
acgagcagct	cgcgcgtgc	caccacggcg	ctcgacccggc	cctccctccgc	gaggcgca	32940
gcgagatcg	tgacgcccgc	ccagccgc	gcgcgcagcc	cgccacgc	cagcaacgc	33000
gcgcggcgca	gcacccgcgag	ccggggagc	ggctgggtgc	gcgcgaactc	cgcgcgcgaa	33060
gtggggacga	acgcggatc	gtacgcgttc	cctgtgcgc	tcaggcata	actgcagccg	33120
tagagaagct	cggcggttcgg	gttgcgtgc	acccgcacccg	ccacccgcgt	cggcgccca	33180
tcggccagcc	ggtcgccttc	ctccacccgc	acgacccatt	cgcccttcgc	gtccgcgatc	33240
gcgcgggtga	cgcgcggcag	cgccgcgtc	ctcggtcgc	gctcgaggac	cacggtctcc	33300
agggcgatcg	cggcgtgc	cctggccggc	aaggcctcg	gcgcacatcca	gtgggtcgtag	33360
gcatcgccgc	ccacgcggcg	gcgcgtcg	cggcgtggaa	agccgaggca	cgcgcggcg	33420
aacaacggat	cgtacgttgc	gggcacagca	tcacccgcgc	catcgctc	cgcgcggcag	33480
tcgcgataga	gcgcgcaggta	ctcgccgc	ttcgccctcg	tcgttctcg	gcgggtggtag	33540
ttcagggtct	cccgacgcgc	atcgagcc	gtgcggtcgt	cgcgcaggcg	cagcaggagg	33600
gccacgatcg	cgcggggatc	gttgtgcgc	agcaggaagc	cggtctcgcc	atcgcgac	33660
cgctcgggaa	tggcgccatc	gwgctgc	gccaccggga	tgccgaggcg	ctgtgcctcg	33720
ctcaacacca	ggcagaaggt	ctcctccatc	gtcgaggggc	gcaggaccag	gtcgatccg	33780
cgaaggatgc	gcggcagggtc	cgtcgccgc	taggcgcgc	gctgcatgtat	gcccgcgc	33840
ttcgcgcgcg	cttcgagcgc	gagatcg	ggaccccagg	cctcgatcgc	gatcgctc	33900
cccgcgcgc	ggcgcgcgc	ctcgatgatc	aggtgcgcgc	ccttgcgc	gccgaagcg	33960
ccgaggaagg	ccacgcgcag	caccggccgc	gcttcggggc	gcggaaagcag	cggcaggctc	34020
gcaatgcgt	gcccgtatgc	gcgcagctc	gcttcgcgc	tcgcgcgc	gcccttgcgc	34080
atgcgcatcg	cgcgtactc	ggacgggcag	aggatcacgt	cggccgcgtc	cacggcata	34140
gacgcgcgc	cgaagcgcga	ttcgaggtag	ccggtcacgg	acgcggcg	cgtcgcccc	34200
acgatcgct	cgcgtcgcc	gcgcaggca	tccacgc	cgccatcg	gccgcgcgc	34260
gcggcgcgc	cgcacggcag	gtcctgcgc	ccggtcatca	tgtttagtgc	ggcgcacagg	34320
tacgagaagt	cgtcgccgt	gatgaccacg	cgtgcgcgc	actacgcgc	gatcgccgc	34380
aggcgcaggc	tgttccagcc	caccaacgac	ttggaaatgc	cgacgtcg	gccccggac	34440
agcagttcgc	ggaaggccgc	ctccacgc	gcttcggca	ccgcgcctc	gtgcccgc	34500
acgcggatgc	cggcgtcat	gctcgtggg	ttcaggcgc	cgagtcgc	gtcgctcg	34560
aggcgctcg	tggtaagtc	ggtccacgc	cgcgccgc	cggtcgc	cagcgtcg	34620
gcgacgtcct	tcgcgcgc	gttcaccgc	gcgcgcgt	gctccat	gccgcgcgc	34680
gagtcgaggc	gatgcgc	ctgcgcgc	cgcgccgc	ccggcattcc	ctcgctcg	34740
gccgcggcgt	tgacgcgc	ggcggcgc	cgcaggggt	tcgcgc	ccacgcgc	34800
acgcccgcga	cgtaggtgg	ccagcgcgc	tcgaggcgc	cgcggttggc	ctcgcgcc	34860
gcgtcgatgc	cggccacccc	gcgaaacgag	gcctcgcc	cgtggcgc	gtagacgtc	34920
tcgcgcgc	cgcgtcgca	gtcccggtc	agcgccgc	tcgagagg	gtttcctcg	34980
ccgtagccgc	ggccatagag	cgagtcga	ggccgcatt	gcacccac	ctcgccgc	35040
accagcatgc	agaagccgac	cgccgtcg	aggcgcggat	actcgccgc	gctggccgc	35100
gcgacgcagg	cggagatggc	atcgacgtc	gctcgcc	cggccgc	gaagatatcg	35160
gcgacgcgc	gcaggctgag	gatcg	cggttggaga	gcggcac	gatccgc	35220
cgcgcatcg	tgtcgcc	acgcagc	ccctcg	aaccggggcc	gacctgc	35280
tccgaattga	gcaccacgaa	gtcgcc	ctgcccggca	ggccgc	gaccgcgc	35340
gcgaagccga	ggttctcg	attgg	acacg	cggc	cgcgcgaa	35400
ctgcccagca	accggcg	gctcgatcg	gtcgctcg	catcg	gacgatcc	35460
tgcgccgc	ccgtgtggc	ggccacgc	gcgaggc	gttcgg	gtcg	35520
ttgaagaccg	ggacgacgac	tagcgc	ggccatcc	cggcgtt	agctggcc	35580
atcgccgc	caggcg	tgctctcg	agagtg	gact	tgcgagg	35640
cgtaggcgcc	tccatgc	cccagg	tgcgcc	ctcg	cgcagc	35700
gctcg	accggagac	tgtgtc	gtcgg	gag	cagcaga	35760
ccgcgtc	cgcctcg	caca	ggcg	gaggaa	gtacatcg	35820

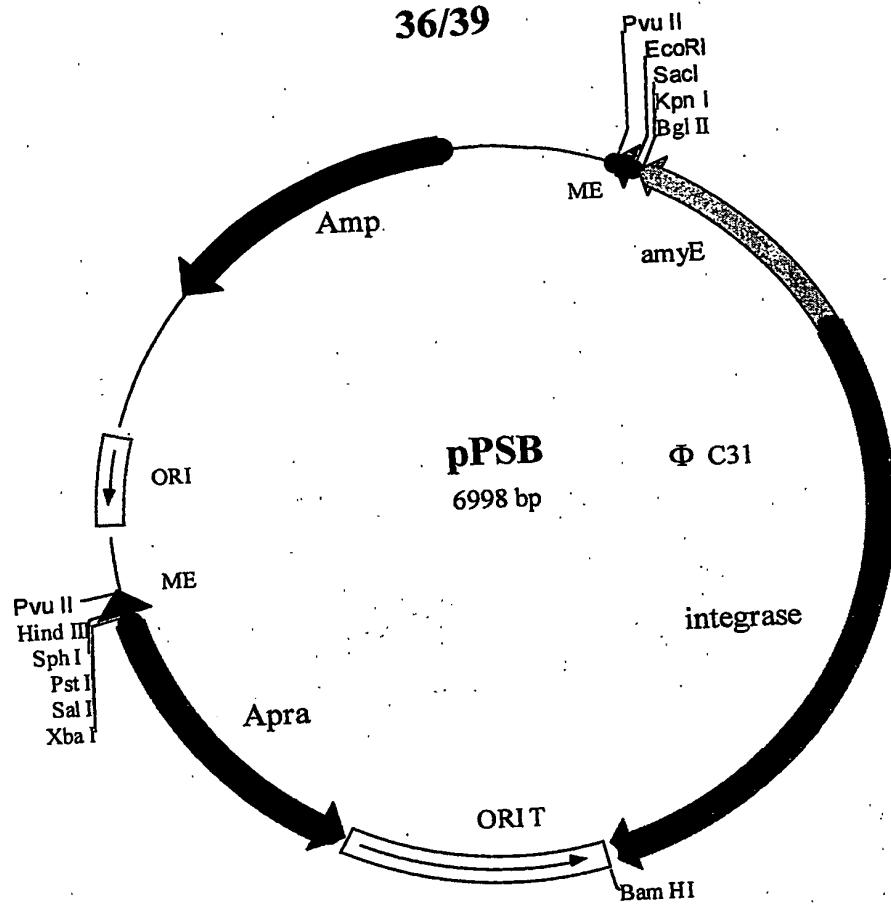
34/39

ccggccgcggg	tttcgcagcc	tccggcgccg	aagcttgcag	caactgcacg	gcgtcgggca	35880
ccgcgtgcag	cggccagatc	gccgaatacg	catcgacgcg	ttgcccgaac	agggccacct	35940
ccgggaaccg	cgccttcagc	acggcgagga	attcgtgctc	gtagagctcg	cgcacgtat	36000
gcgggttgcg	gtagtcgcgc	tggtccgagt	acacctcgcg	gttcggcgtc	gagaccagca	36060
gcaggccgccc	gggcgcccagc	acgcgcttcg	cttcgtcgag	caggcgctcg	ggatcgggaa	36120
tgtgctcgag	cgtctcgaag	gagacgagga	gatcgatgt	tgcgtcatcg	cacggcagcg	36180
actcgcagcg	gccctcgacg	tactgcaggt	tctgcgcggc	accatagcgg	cggcgcgcct	36240
gcgcgatgtt	ctcggcgccgg	acatccgcgc	ccaccacggg	cttcgcgcgc	gtcggcagca	36300
gcgcggagcc	gtagccctcc	ccgcaggcga	tgtcgaggac	gcccgcgcct	cggcggagcg	36360
gcaacgcaaa	gtggtagcg	tgccagtgct	cgtaccagat	ctcgccctcg	aagccgggct	36420
ggaaacgttc	gtgttccatc	gggagagtat	aggtgtatgt	gcaaccgcgg	tccggagagg	36480
gcgggttgcg	tgatcttatt	cgagatcggc	tgcctacggaa	gcaccggcaa	cggctgcggc	36540
ggcgccaccg	accatccggt	gcccgcgcgg	agcagggttcg	cgctggcgag	cgtgttgagg	36600
ccgtccatca	ggcgcaccgt	gatgcggcca	tgcacatttc	tgaacaccat	gtcgagcttg	36660
ccgtcgccat	cgaagtgcgag	cagttgcgtg	acggtccaggc	ccgcggccgc	cggcagcagc	36720
tccgcccgc	tgaggatggc	ggtgccgtt	atcaggcgca	cgtgcacacg	gccgtcggtg	36780
tggcggaaaca	cgtatgtcacc	gcccggctcg	cgttcatgt	cgcacgttg	gctcacgtac	36840
cagccgggtc	ccgcggccag	gagctcgaa	ccggcgccga	acgcgtgtcc	gttcatgtg	36900
aagagatgcg	cgcggccatc	ggtgtggcgg	aagaacatgt	ccgccttgcc	gtcggcgctc	36960
acgtcgccca	ggtgcgtcac	cgtccagccg	ctggcgccggc	agaggaagcc	cgcggccggcg	37020
gtgatggtgg	tgcgttcat	caggtagatg	tagccggcggc	cgtcagcgtg	gatgaagacg	37080
atgtcgcccc	tgcgtcgccc	gttgaggtcc	ccgggtggcca	cgactctcca	gcccgtcgcc	37140
ggcccccagca	gctgctggct	gccgatgt	gccgtgccc	ccatcagcca	caggtgcgag	37200
cggccgtcgg	cgttgcgcag	caggagatcc	gccttgcgt	cgcgttcat	gtcggcggtg	37260
cggtcgtatgc	tccagccgg	gcccggccgg	agcagtcct	tgccggcggt	caccgtgagg	37320
ccgttcatcg	tgtacacgta	cacgcggcccg	tccgtgtgg	gggatctct	agagtcgacc	37380
tgcaggcatg	caagctttag	tattctatag	tctcacctaa	atagcttggc	gtaatcatgg	37440
tcatagctgt	ttccctgtgt	aaattttat	ccgctcacaa	ttcacacaac	atacgagccg	37500
gaagcat						37507

35/39



36/39



Tn <amyE-att-int-oriT-AprA>
4424 bp

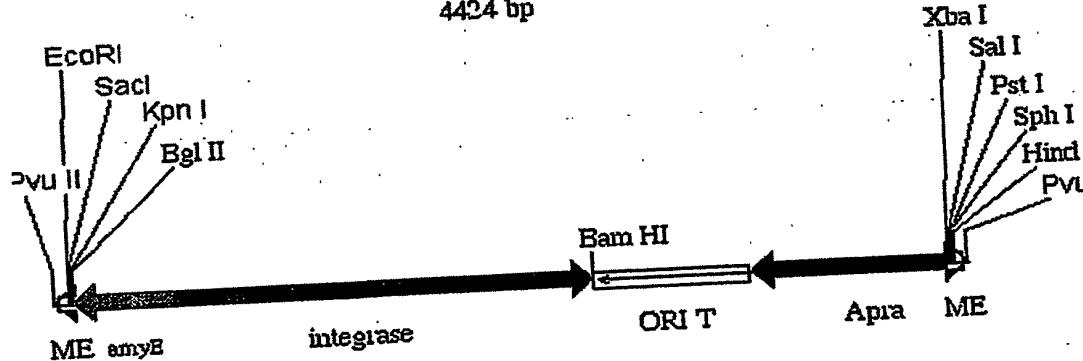
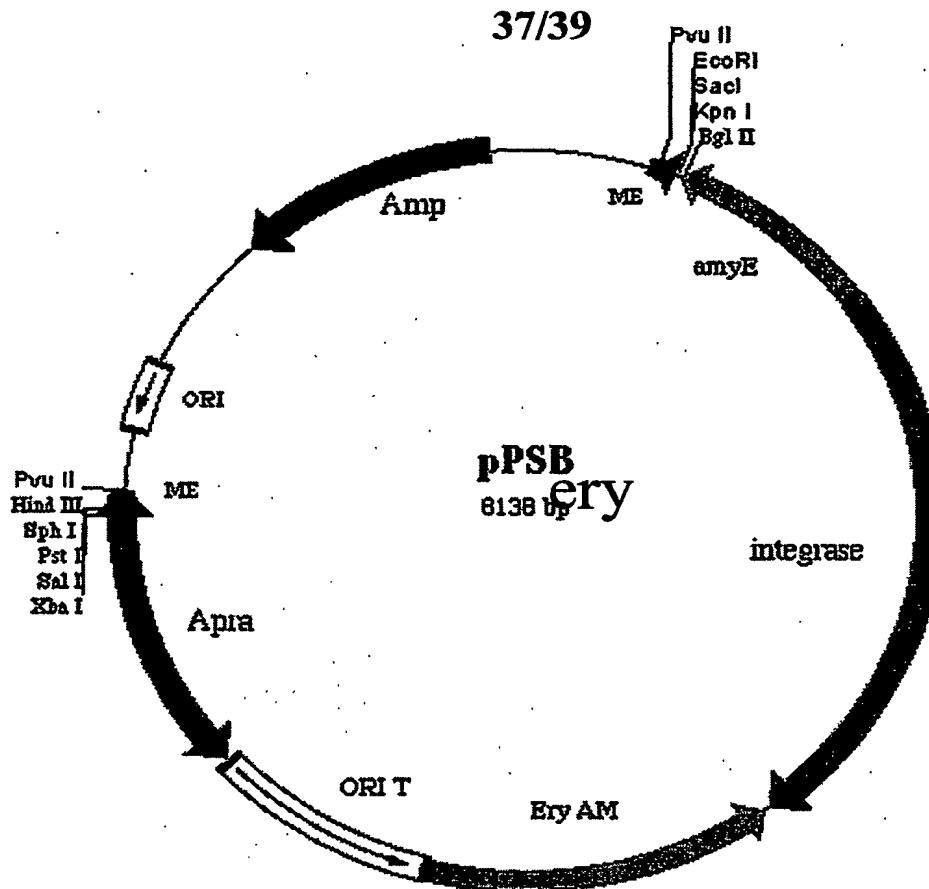


Figure 7



Tn<AmyE-att-int-ery-oriT-Apra>

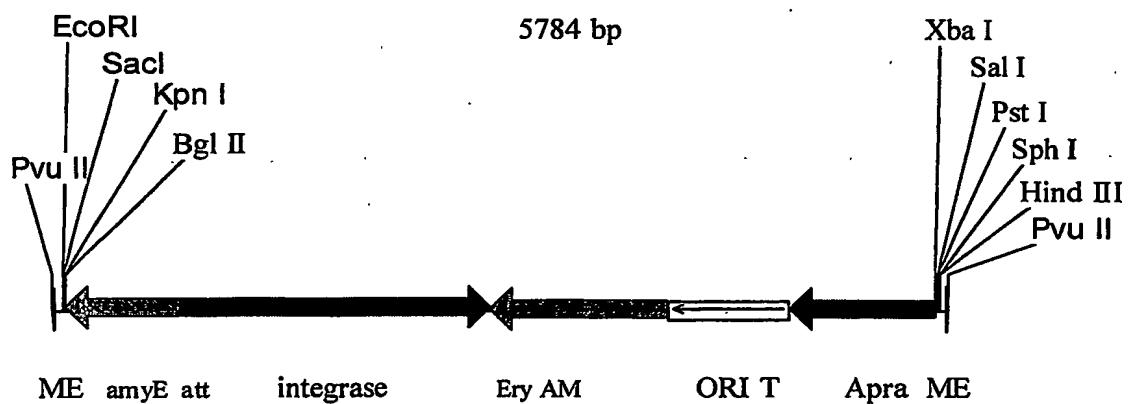
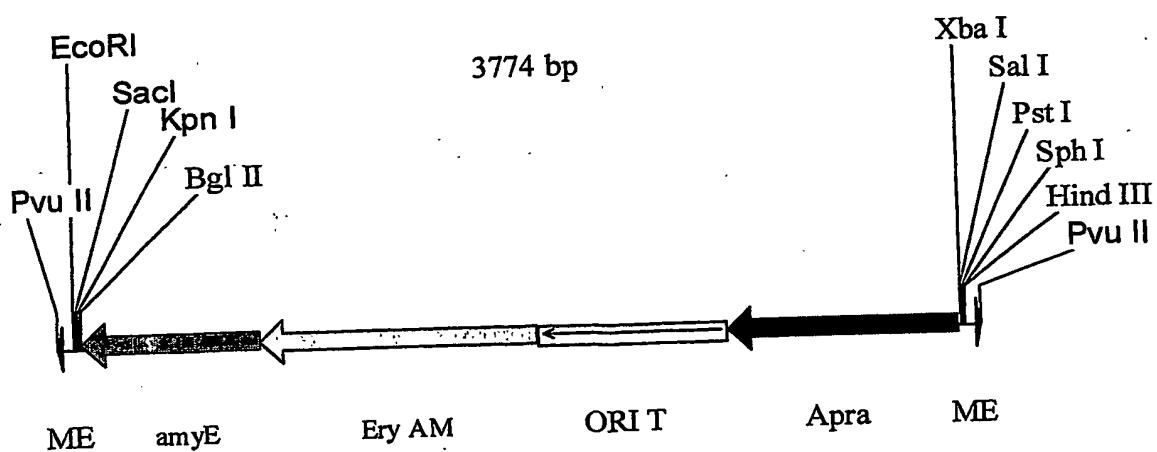
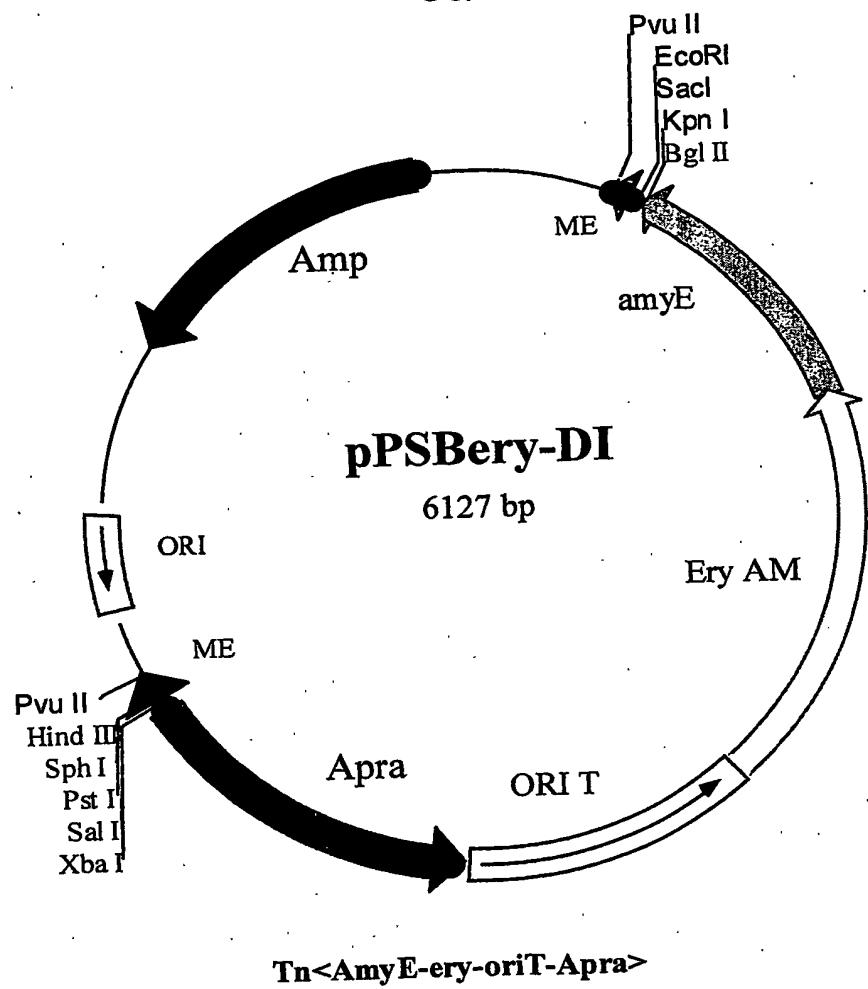
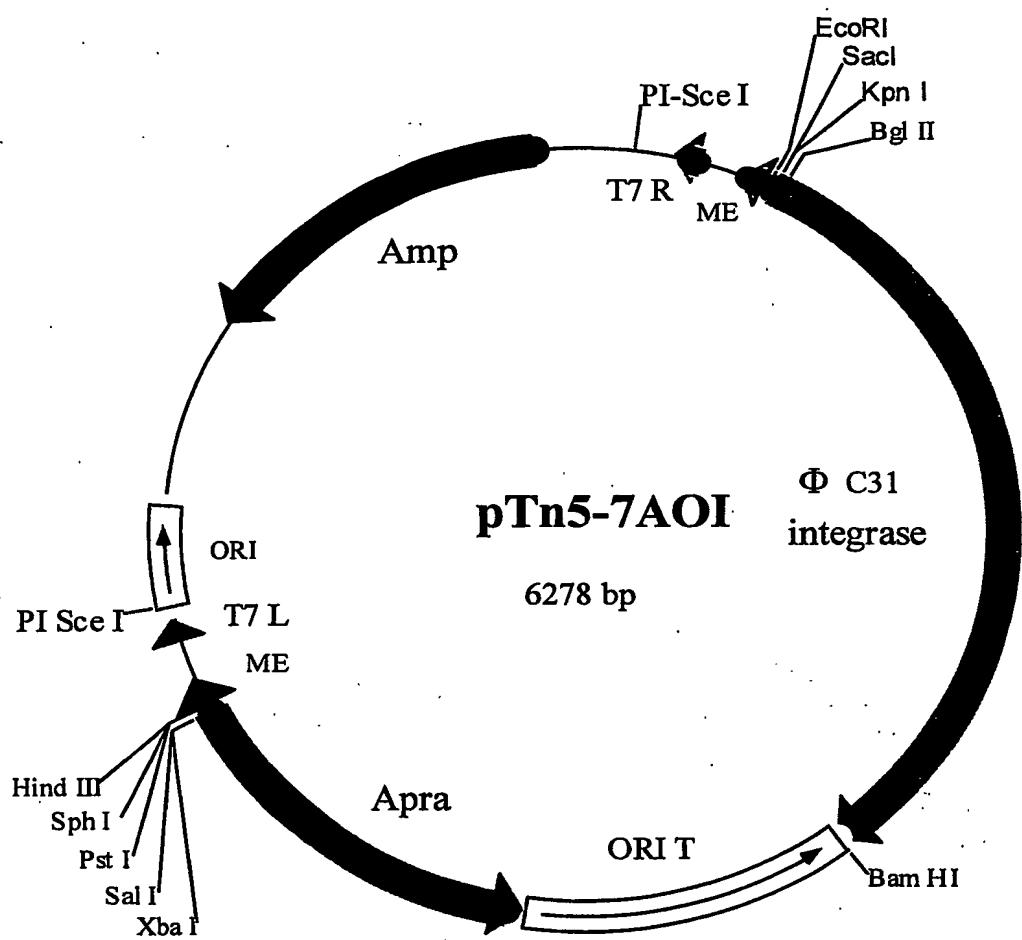


Figure 8

38/39

**Figure 9**

39/39

**Figure 10**

SEQUENCE LISTING

<110> LIBRAGEN

<120> Method for the expression of unknown environmental DNA
into adapted host cells.

<130> B0149WO

<140>

<141>

<160> 16

<170> PatentIn Ver. 2.1

<210> 1

<211> 37500

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: DNA sequence
of clone FS3-124.

<400> 1

gatcggtgcg ggcctttcg ctattacgcc agctggcgaa agggggatgt gctgcaaggc 60
gattaagggtt ggttaacgcca gggttttccc agtcacgacg ttgtaaaacg acggccagt 120
aattgtataa cgactcaacta tagggcgaat tcgagctcg taccggggta tcccacgtac 180
caccgagctca tcttggaaagag cccgctccag acacccgagg aatggaaagcg cggcgtcgt 240
agaccgtaca cgcagaagcg tctcgtcgcc ttctgtact ggcttgcac gacgatgaag 300
ctggatctca ccaggatgtt cggcccgaggc agctcgatgg ggccatcgcc cgcgatcatg 360
ctcgcgattc gctatcccgc acgattcgcg tggaccgtgt cgtgggtcgg cgtccacgt 420
cccgccgact ctccgcaatt cacatcgctg tacgagctgg tgcacggccg gcccgactt 480
aagggtgccgt tcgagaacgg cacgcccgtc tggatcatt ttatgtacgt ctggtaact 540
cggcagcatc cggagcggc catcggttc atcacttct cgaacggcaa gaacgactcg 600
cgatcggtt ggcgccaggc cgtcgaaattc ctgaagacgc tgcaggagac caggcaaccg 660
ggcgaacggg agatgcccgt cgacccctaga acgaaccaga gcctgccc gttcagccgt 720
cacctgttcg tctggggcca ggagggacac ggccagcg cgaagatgcc ggaagggtggc 780
ggcgaacggg agatgcccgt cgacccctaga acgaaccaga gcctgccc gttcagccgt 840
tgcacgctcg acgacgatcc gggagacggg tcggattcga gcccgtccc ggcggccag 900
atcaactctt acgtcacctg gcagcgtac tcgatcgctg acgaacccgg ccgatggagc 960
gtcgcatca agctggcgaa tcgaccggc cggtcgccc cgcagtcga cgtcacgccc 1020
cggcggtgc agcagttaaa gctcaagccc ggcgatcagg tgacatggac gaactcccg 1080
ggcaacaccg tcgtgcagcg cggcgaagcg gtcgcggacc ggtgggtct cgtcacgctc 1140
ccgcagatgc aggtctcaac ggggtggaaac aggatctgg tgagcaggaa ctaggccggc 1200
ggcgtccctcg acgagccggc gcatcgccgc cagttgtatgt gcccacgaaa gctggctgag 1260
gtccgcgcgc gcccgtatgt cccgtcgcc ctgcggccggc tttgccagca aatccagcac 1320
agccgccccgg aacgcctcga ctgattccgc cacggagcac gcccggccgtg cgatgggtgg 1380

aagtccccct gcgacggccg gggcacgac gttcgggagc cccgcggcaa tcgcttcgag 1380
cacttttttc tgcacgcctc tcgcgcgcgc aatcgccgcg accgcgcgcg ccgcgcgcgc 1440
gagatactcg cgtaccccg gacccgcgcg ggtcacttca atcgtggat cggccgcgc 1500
tgcccgccacc gcggccgtcg gttccatgcc cacgagcatc agccgcgcgt cccggccggc 1560
gcgaaccacg gacggccaga cctctcgccg gaaccagagc gccgcgcgcg ctttgggctc 1620
gtagccgaac acgcccaga agacgaccgg cgcttcggcg gccgggtggc ctttcggcc 1680
aaacatctcg acgtcgatgc ctttcggcac gaccgacgcg gtgaggccgc cctgtacttc 1740
ttcgagcgcg aggccgcgcg gtcgcgtcac gaccgtggc gccgcgcgcg gccgcgcgc 1800
gactttctcg aagcggagca agcggccggc ctcgcggccg aacagccagc ctttcggcca 1860
cgccgcgcgc gacgcgcagcg cgcgcattt ctcggagtcg acgtccacca tgtccaggac 1920
gaaggggatc cccgcgcaga tcggctcgca ggcgtagcg gccatcccgc tgcaagtacgc 1980
gagcaccgcg tcggccgcg tcgcggcaat ccgtcggtcg agcacctggt agatctccgg 2040
cgaatgcgcg aggacgtgcg tgaggggtcg atcgcggccg agcgcgcagac cgctcgcgat 2100
gagattccgg cccgcggccg cgcgcacgcgac gtcggctcgac gccgtcacgc cggcgagatc 2160
cgctcgatgc gaccacttc ctcgcgtcg cgcgcagcg acgaggtgca ctcgcggccg 2220
tcgtgcgagc gtgtgaatga gatgaaacgc gcaaatgcgta tcgccccgt tcggggcgta 2280
ggcagccgg tttgtgagca gaaggactct caccgtgcgg catccgatcg ctctcgatg 2340
acgcggccgcg gtacgcgcgc ggcgcacgcg cggcgccggaa tcggcgccgt cacgaccgc 2400
cccgccgcgc gcaatc acggccgcgc tcagccatgc ccacggccgc gtcgcgcgc 2460
caggtgcgg cgcgcgcgcg caccatccgc ggtgtgcctg gtcgcgtcacg gatcgccgc 2520
gtgaggcgcc tcgtgcgcgt cgtgtccgtt ccgcgtgggaa ttggacgcgc ggcgcgcgc 2580
agacatgcg gtcgcaggcg gacccagccg agatggcgtc cgggtccccatc gtacacatgc 2640
tcgtcgagct tcgcgcgcgt ctgcggaaatggtgcgaa aggcgcaccgt gaccgacggg 2700
tcgcagtgtg caagcgcgcg gccgaggaaa gccgcgcgcgaa ggtattgtcc gatgacgcgc 2760
ggatgagcg agagccactg ctcgcgcgcgt tcgagggccc gtcggcgcc gaggaaaggc 2820
cgccgcacgg agtaggagag caaggcgccg aggacgcgc ccagcgcgc gccacgggg 2880
agtccctgc cggcgccctt gaggctcgcg cccactgcgc cgcgcgcgc gctcgccgacc 2940
gggtccgta tcgtgtgcg agcgcgtcg agacgcgcgc cacttcgcgc atccgcgcgc 3000
cgaacgataa ctcttctcg atccggcgcc ggcgcgcgc gtcgcgcgc 3060
cggttgcgc catcgccctgc tcgatcgatcc cgcgcgcgc tgacggccgc cccggctgaa 3120
tgaccaggcc gtcgcgtccgc ttgcacacgc gtcgcgcgcgt gccccggcc gcggtcgatc 3180
cgatgggtgt ctcgaacgcgc atggcttcgc gacggcggtt gggcgccct tcgttagtcg 3240
acgactgcac gacgggtcc agccgtgtt ggcgcgcgc gacgtcgccgac gtgtggccgc 3300
ggaaacgcgc gttccggcc cccgcgcgc gggcgccgtt ttcctcgagc ggacgcgcgc 3360
ggctgcgcgc gccggcgatg acgagacggc ggcgcgcgc cttcgccatgc agggcgccac 3420
aggcttcgcat gagcaagtgc aatcggttc gtcgcgcgcgt acggccaaacc ggcgcgcgc 3480
cgaagtgcgc ggcgcgcgc cccgcgcgc cccgcgcgc ggcgcgcgc cgtggatcg 3540
gtcgaaagc atgggtgtcg atgcatttcg gatcgatgc gacccgcgc ggcgcgcgc 3600
cttggccag gagctccgc cgaatctgtt cggacacgcg aatgacgcgc gggaaaggcgg 3660
cgagaatccg gcgatcgagc gggtagtgcg cggatcgatcc ggcgcgcgc ggcgcgcgc 3720
atccatggcc gtcggccagc ggaatcgcgc gtcgcgcgc tcccgaggac agggcgagga 3780
gatcggtttt gtaatcgatgc gtcgcgcgc ttcgtatcgatcc cggatcgatcc acgagcgccgc 3840
ggagcttcgc ccagattccg gtcgcgcgc aatgatcgatcc cggatcgatcc acgagcgccgc 3900
tgccggccgc ttgcgcgcgc tcgtccatgc cgaacacccgg atcgatcgatcc tcccgaggac 3960
acgagacggc gatcgatcgatcc ttggagcgat cggatcgatcc cggatcgatcc acgagcgccgc 4020
tctccggatcc gccggccgcg cccgcgcgc tcccgaggac gacgcgcgc acgggacgcgc 4080
cgattcccg cctaggcgcc gtcgtatggca gatggcccc cccggccgc gacgcgcgc 4140
ccccggccgc tatcgatcgatcc gccggccgc acgagcgact gtcgcgcgc gacgagtc 4200

actgcgccgc gccggcgccg gatcatcgcg gctgcgcgca agccggcgaa accggcaagc 4260
agcaggagcg cggtaccgac aatccaccac cggccgcgct gccacagcgc gaggcccagc 4320
ggaatcgaga cgagcgccag cagatcgagg accgggatga tcacgctggg cagatcccg 4380
atggcgagcc ggtgacggaa gtcacgcgc aggttgtcgc gggcgccca cagctcgccg 4440
aaaaacacgg cgcgaaagcga tcggggatcg ccctggtgca cgctgcggag cggcgggtcc 4500
gacaggatgc ggccgcgcgc cggcgcagc cgctggcaga tatccacgtc ttcacaggtc 4560
tcgaggctgg tatcgaagcc accggcgccg tcgaaggct cgcgttac cggcaagattg 4620
ccgctcggtt gccagttccac gtcttcggtg ctctgcccac gacggcgaaa cgagtcgtac 4680
gcgcgctgca cccagttgcc gtcaccggga gcttcgtact gcgccgcggc agcgacgacc 4740
cccggggctt ggagcagatc gacggcgacc accggccagc ggggtcgat ctcatgtatcg 4800
gctgtcaacga acgcccagcac gtcgcccetcc gccgaccggg cggcgcgggt gcgaaagcgtc 4860
gccacgttca cgtcgggcag cgacagcact cgccggatc cgctctgc cacggcacgc 4920
gagtcgtcg tcgaccctt gtccgcccacg acgatctcg accagccagc cggcgccctcg 4980
gtgcgtatgaa tcgacgcccag gcaacggctg aggtgcgcgg cggcgttctt gacggggatg 5040
acgaacgaca cgccgcggcgc cacggcaatt ttgcgggtca tcgatttcgt agcgcggccg 5100
gctttcgtag cgccggcccg gttgggggccc gcgagaactc gaacgcgccc agatccggcg 5160
ccctgtctt cggccgcccga tcgaaaatcga gcccgcctg cgaatctcg atgcccgggt 5220
cgatcgctc cctggcggtt gccgacaggt gcagatcgca cgaggccgcg tttgcgaacc 5280
actctgcggc ggcgttcgtg acgttgcgt tcgcctccgc ctaccgcgc ttacggcgca 5340
ggatctgccc gtcgtcagg ttgttgcgg tgagccgtt cgtcacggga aaccggacgc 5400
cgatcggttca cggcgtcgcc gtgcccagcg tgaagacgtt gttgtctcg atccggaaatg 5460
cggtggacgc gttggcttcg atcgcttcgt cggccattc gtcagggtt cagatcacgt 5520
tgttccgcac gatgccgcgc tggtgatcga tctgttctc gcccgcgcgt gcgtaggccg 5580
tgatccccgc ggcaggccg atggcgatgc ccctgaacga atcgacgatg acgttccgtt 5640
cgatgagcgt tcctgcgag ttgcggccagg cgaggatcg cggccggat cggcagccgc 5700
ccgattgcgg gccgcaatc cgcaggaaga cgttgcgcgc gatgaccac ccccgccgcg 5760
ccaggatgtc cacgcccgtc gtgtaatcgg acggcgcgt cgtctgtac tcgaagcgcg 5820
agcacccac gaggccatcg tccgcgaagc gtccgtcagt gcccacgcgtt ctttgagca 5880
attgctgtcc cgcacatcgat agctggacgt tgcgttccgtt ggctcgggac ggcacaagct 5940
cgccgcgcac ctgaatcgcc tggtacccga cgtggccgac cgtcagatcc gcaatcgta 6000
cgtcatgtggc gccaaccgag agcgcacgc cgcgtcgatc gcccgtcatg cggccgcgc 6060
gaatgacgac gcggtcccgat tcgcccggcgg ggcgtccggag cacggcgcgg gggacgttga 6120
tgctggccat ccggtcgagc cggtaactcg cgttctcgat cagaatcgatg gtgcggcgt 6180
tgacgcgttc gagcgcctgc agcagtcgg acgtccgcgc cacggatcgacg acctgtttcg 6240
tcgcccggcgc cggccatccg cagaagtgcg gcacccggcgc ggcggcctgg cggacggca 6300
gaagcgccac gagtgcgcgc gccacgtat ggcgagtcat gcccgcaccc tgctgcgtc 6360
gtgttcgcgc aaccagagct cgaggatgag caggatitga aggaggtatc cgcgcgactg 6420
tctccgtga cgcgtgtcg cgtatgtatgc ggcacacgtc tcggggcgca ggaacatccg 6480
cgtgcgcgc cccggccgc acacgtcg ctcgacggct tcgagcgtc cgtctgcag 6540
ccagcgatcg aagtgcgtat agtgcacgtt cccaggcgcg ttcaaggcgt tcagcagggt 6600
gttgcacccgt tcgaacaccc tctccggcag cggacccggc tcgcccgcgc cggccgtt 6660
cgagttccgc accttgagca ggcgggggtt gcccgcctgc gtatgcgcgc gatgaagcgc 6720
cgtgtcgatcg cggccatccg gatgcgcgc cagcaggacc tcgaggtact cgcgtatccat 6780
gaacggcgc cgcacgttca cgcacgtcg gaagagatcg agcggaggaa tcgtgaagcgc 6840
ccgatgtatgc tcacgaagat agagatagct gctgagctcc gcggttgaga gctccacgg 6900
gtcgaggagc gcggtcgatcg agtgcggccgc gcccgcctcg gagaggcgcg cagcgtctgc 6960
tgtgaagagc gtgtccagcg gcaggccgtt gctgatgtatg ttggcgccgc acgtcatgt 7020
gtcgatgaac tcgcccggc cggtcatcgca gacacgtgg cgcgtcggtt ggagcggcca 7080

cgccaggctc gtcttggcga gctctccgcc gtggccgcgc agcaggacct cgataccgga 7140
gtccccgagg aaatcgatcg ccagcatctc ggtcagtccg tggctcagat acatgccatc 7200
ggtgagcgag accatgcgcg cctgggtcgg cagaaaagtcc ttgaggtacc ggtcgtcgag 7260
ctcgaagaag cggtggtcgg taccgggttag cttcgcacagc ttttgcgcga tcacctgatc 7320
ggcgcagccg ggcgcacccc gcgtgtacgt gatgagcccg ccaacaccgt tcaccgcact 7380
cagcacggcg cggctgtcga ggccgcgcga gagcgcacagc cgcgaacgcgt gcctgcccga 7440
gagcgcacttc tcgaccgcgc gcctgaacgt cgtgtgcacc gctggacgt acgcgtcctt 7500
cgctcgatcg ggtccgcgg agagatcgcc cggctcgaaa taccgggaca cgctcgccgc 7560
gcctgacggc cagtcgtacg cgagcacggc gcccgggtcg aggagcgtga cgcccgccgc 7620
cagggtcttg tcgcccagga cgaagcccg cgtgacgtaa tcggcggccg cccggaggct 7680
gagagccgggt ttgcgcgcgc aggtcctgtag cacggaagcc agatccgatc cgaacgcacag 7740
tccggccgcg tcgaccgcgc agtagatcggt atacgagccg aacggatccg ttgcaagcag 7800
caactcgctc cgcgcggat cgatcacggc gagcggaaaac gcgcgcgtcga gcccggccgc 7860
catggcaggc cccagccgtt cgtacacccaa agccaggaga tcggcgcgcg tcccgttcgc 7920
cgaacccaaac tcgctccggc ggcgcgttc gtttagaga tcgcccgcga agatcaccag 7980
cgcccggtcg acgtcgccgc cggccgcac ctcgtcgagg aagccgagcg tgccgagccc 8040
gaccggccat cgggtccgacg gatctcgata gacgaccgtg cggcgcggcc tgcccgactt 8100
cgccgcgtacg gcccggctgg cgcggctgaa ggcggtagat gccggctcat cgcggcaaaac 8160
gacgcccggagg aggcgttca tcagacgacc cccgcggcag acacgcgcga cggaaacaaga 8220
ggtgtcgtct gagcggccgg aaccggccgc gcggcgtcgc cgcgcaccat cccgtttagc 8280
gcaatcgaca agccgaggaa gtgccagagg atctcggtgt actgaaaagat gacgaacgtt 8340
ccgcccacgc agaatcgaaat gaggcgggtc tggacggcca gcgcgtacgg cgccagctcg 8400
gcccaggccg gatccgcgtt cgagacccgg cggaccctgc ggcaggcgat gagcgcgcagg 8460
aggagagtga gcaggaacgc cgcgaggccc acgaacccaa gtcagccac gagaccgaac 8520
caacgcgtgt gaaccgcgcg ccccttgcgc caggcgcgtt cgcacgtgtc gtatcggttg 8580
taggcggccg tgaaggcatt gtgaccgacg cctgtaaatcg gccggtcatc ggccatgcgg 8640
agcgcgaccc tccagaagtg cagccggctc tggccgatc gtcgatcgcc aaccggccgt 8700
tcgctcgctg cgtatcgatc gtcggacgag ctgatcgagc tcgatcgatc ccagaactct 8760
tccgtcatca cgagcgacat cgcgacgaag accggccccc cggccacgcg gacgagcatc 8820
ttgcgccttc agtacaggaa atagacgagg ccaagcacgg cgagcgttaga gaagccgccc 8880
cgggagtagc tgctgtatgcc gcgcatacgac accccgatga gcagcacgc cagcgcggccac 8940
ttggcccatt tcctcgcttc cgtctgcatc agcgtgacga tgagcggcac gacgatgaac 9000
atgcccacgg cgacgcgtt gttgtcaccg agcgagttaga cgtcgtttaga gttcttgatg 9060
ccggaaatga acaggagctg accccagccc tgcttggcgg ctcgaacccc gagggagagc 9120
gagatcacga ggaagacgag cgcacccgc ttgacgtccg tcgtgaggac tgccgacgg 9180
tacgtcatca cgggtggactt catgaagtgc atcatgttagc cccacgcgaa gtcctggtag 9240
ggcgcaccca gcgtcgacga caggctgtgg accccgaaaga aggcgaggag cgcgcgcgc 9300
gcgtcgatcc gcagcgccctg cccggacagg aacgcgatcg cgacgtgtc catgcccgc 9360
aggagcgaga tggccatgtg gatgagaaaag tcgctccaca cccagagctc cggccggaaag 9420
tacgcgtatca acaggttagaa cagaatcgacg tagaacggcc cgcgcacgc gtggaaaggcg 9480
ccgaaggcga ggagggcggag gacgaaagcg gtcgaaagca ttgcccgttg ttacctgtta 9540
tcgggtcccgaa tggccatgtg cggctgcgtc cgggtctccct ctacattgaa caccttggtag 9600
agggccggaa tcgaggtgaa catgagcacc acgaaaagaa tcgcccacac caccatgttag 9660
gcaagaacc cccgcgtctt gtagagcttc tccggattct gcacggggct gttcggcagc 9720
atcccgatgt gcaggttagta ggccaaatcg cccgcgcgcac cggccgcgaa caggatcgac 9780
tcgaggtgat agcggacgtat gaagatgcgc cgcacggagg cgcacgcgtt ggcgttagaaac 9840
accatcgatca cgcgcacccgc ctccctcgatc tagaaggcaa acgacttccg gtacgaggcc 9900
gcgcacccgcgtatc gtggccggaaac tccgcgtacc gtcgttcgc catgaaacaac 9960

gcgccccacca tccagttacga gatcacgagc gacatcgccg gcacgcgatc ctgaatcagc 10020
gggaaccacgc cgaggaggag ccggacggcg ttgttcgccc actcgctcag cacgtcgagg 10080
tacggccact ctttcgtgcg aatgggcggc acgttgtacg tgacgcccag cacccagagc 10140
gccaggcccg acagcgcgaa gtatctgttc acgagcagcg cgagcacgaa tccggccacg 10200
cccacgagga tccactcggt ttagcccgcc gccggcttga tcttcccgga gggcaccggc 10260
cggtgccgtt tctccggatg cagcagatct ctccggccgt ccaggagctc gttcagcagc 10320
tagttgtcg aggcatgag gcaggtcgcg gcgagcgcga ggcgcgacgg 10380
gtccagccga agagctgcgg ttcgtagaag aacgcccagca gcacgcgag cagcatgaac 10440
gcgttcttga accagtgtatc gatccgcgcg atctggacgt acggccaaat ccgcgaaccg 10500
gagctagcgg ctgacgacat ctttgagctc ctgcggccgc cgatcgcgaa accatccgcg 10560
gccggctgtg acgcgcgtggc ggccggcgat gagggcgcag gagagtctg cggccatcgc 10620
tccttcctccg tccacgtccg gtcgatcccc gacatacagg acctcccgag gtccgagtga 10680
ccacatctca cacgcccacat ggaagcccg cggtgcggc ttcaggcggt tcaacgttgc 10740
ggcggctgtg cagagcgcga gcaaaaaatg ctgcgcgtg ccgagcgcgg cgagcttgc 10800
ctcggcgccg tagtccgaca gacgcgcgag cctgagcccc gccgcgcgaa gcgagtcgag 10860
cgcgccgagc aacccgggc gccggccacca ccgcagatac ttgagcggcc ggcgcaccat 10920
ccattcgttc accgtcgccg ccacgaccc tttgtccagc ccgagcgctt ggcgcgtgc 10980
cgcgatctgg cgccgcgcga gcccgcgtc ggccgcgccc aggcgcggaa gatcctcatg 11040
cgccccggcg tactcccgca cgatgcgcgc cgtgtccact ccgcgcgc accggaacgt 11100
cacgagcggc gtggcggcga gtcgcgtgc catggcgccg cgcaacggac cttgcggta 11160
cagcgtgcgg tccacgtcga acagcgcgc tttgaccccg cgcaacggct tcaactctgc 11220
aggcccagca tcccgagaat cgccgacacc tttgaccccg cgcaaggaa cacgtcacga 11280
acccacacgg cagccagccc cgccaaacagc accggcgctg ccgcgcgc tgccgtccgc 11340
cagctccgcg gccaggcgccg cgtgtcgatc gcccgcgcg cgtagcagag cagaatcggg 11400
acgagcggca ggtgataccg gggatgtccg aagacgatcg cgtgaatgcc gctcatgaac 11460
acgatcaggg tcaggagcag caggtgcgcg cgccagtta ccggccgcg gaacgagatg 11520
ccgatgacgc cgagcaccat gaccgcgcg tacgatccca tggatggacac cgccccggcg 11580
aagaacagcc accgcggcg acggtagaac ccattccga ccgcgcgcgat gaagtcgcgc 11640
tccagccccc agaagtctgc gaacttcagc accgaacggc gaagcgttgt tcccgatgg 11700
gccttcatgt agtcgagcgc ctggcgctgc gcccacttct ctttcgtgcc ctccgtccag 11760
cggtgcgtt ccggcgccgc aggccgcatac gtgcgcgacc attcctgtc gcccgtgagc 11820
gagatcgcgtt cccacatgcg atcctcgccg gtgtacccat catcaggatc 11880
agcccgccga gctgtccac gaccgtgaac gagcgttgca ggagcgtgtt ggcgcacgc 11940
cacggccga cgatcgccgc gtatccggcg aagagaagca cggccacgcg cagacgcgc 12000
cgaagcggga tgcccgagacc gacgagcgcg ccggccatga ggatcgcgcgtatccgc 12060
atgatcgagc gaacgagcgc ggacgcgcgc agcgcgcacgc ccgtcgccgag cgatccgc 12120
ggccgagtgcc gccctgatgg gcgatcgatg agcgtcagac atccaaagag caacaaaaga 12180
aggagtgtga tgaagagcgt ctggacagc accagcacgc cggagaacag cagcgggg 12240
tagaacgcga acgcccgcgc ggcaatgagg ccggcgccgtg cgctgaacaa cttgcgcgc 12300
atgagataca cgagccagac accgagcagg ctcaacggaa cttgcgcgc ccggaccgc 12360
gtcagcgtt ccgcgcgtt gaccgcgcg acggccgcg cgaacgcggg aaagagcggc 12420
gcccggatgg acgtcggttc ccccgcccc cacgcgcgacc cttcccgcc gacgacgtt 12480
ctggcaagcg tgcgttagtg ctgcgtcg cgaatctcgaa gcgcgcacgc cttgaacccc 12540
caaatcaggc cgagccggag gaccaggccc agcgtgagga tgacgagaat tgcccgccg 12600
tcccaatggg tgcgtcgcc ctgcgtcgcc ccggggccgca cgattggcca gggggccgac 12660
gatgggagtg ggggagcctg catcattcgca taaaacaagc agttacggc aaacctgtgc 12720
caccggggaa tcagcaaaga ctctgcacac gggtccgcgc gcccgcacgc ccggccgc 12780
ttttccggcc ggaagtcggc ccccaaggcg accggctgg gcacagccca gtagaaatct 12840

cattctggga tgactttgt coatgatcca tgaccattc gaggggcctt cgtgacgccc 12900
 tggggatggc ataagcttgg cacctgcgt ggcgacgtgt cgatgatcag cggccgtgag 12960
 gagctccaga aggcctatcg ggacgaccgg gtcgcccgtg aatacgtgc gcggccgttc 13020
 cagtcaccgc tggggcgt ctcattcg cgtcagatcg gggtcgtgcg agagctcg 13080
 cgcgcgcagg gcatccggcg cgcggcggag atcgaccgg gtcgcgcg gctgaccgtg 13140
 gacatcgcc cgctccctcg cgcgtgacg ctgcgtgcg cgagcgcaca gatgcttcac 13200
 gaagcgccg agcggtcg cgaccggacg ctgcgcgc gggccaaatct cgtgcaggcc 13260
 gacgcgttgc gtcgtccgt cagcggtcg ttgcgtctcg tctacaccctt caggctcg 13320
 cgcgcatttc agcgccgca tcggctgcgat ctctaccggc agattgcgac gattctcg 13380
 cccggcggt ggctcgctt cgatgcggtg aatgagattt tgctgcgcg gcttcgcgc 13440
 aacgcgaagc cggcgagta cgagcaatc gatgcgtgc tcacaccggc cggctacgc 13500
 gacgagctgg ccgaggccgg gttcacgtg tggtcgctca cgggtgcgca tcggccgttc 13560
 cccacgctca tgaagtgcga gatgtatctc ggcggaggt caacgggtcg cgcgcgcgc 13620
 gcgatggagg tgatcgatcg gtcggccgc gaaccgctcg agtggatcg cgtatgcgc 13680
 cgcgggtgac ctactggacg ggtacctggg atccgatcaa agaagcgatt tgaaggaag 13740
 tgaacgcctt ccgcaccggg tgcggggcgac accgcacggt cgtgcgttc tcggccggcc 13800
 agcgcagccg ttcgtgccc aaagatcgcg tgctcggtt gtccggccgc gcctggatcg 13860
 cccttcgggc cgccggccgc gtgcgtcg cgcgcggcga cgtgcacgc acgttcgc 13920
 gcgggttctc ctgcatctc ctgaggcgat tggggccgcg gccgggtctcg ctaccggcc 13980
 tcaaggaaga tggcgctgcc gatctggccg gggcgatcgc gcacgtggcc gtcgaagtgc 14040
 gccgcgtcg cgacacgtgg atcgccgcg gtagtctcc cggccgcgtg cacctcattc 14100
 atcccgccgt ggacctcgac tggatcgcc cgtgcgggtt ctcgtcgag agattcacgc 14160
 tgctttcgcc cagcacgcg tccgatccgg ctgagatcga cccgcgcgcg attccgctcg 14220
 tcgtcgagct ggcgcggctc cgtccggaga tcgacgtgt ggtgcgtgg cccgggtgg 14280
 cgcacgtcg cgcctccagg cgcgcgtcg acgcctccg tccgcacccg aacttcatcg 14340
 tctcgatcgaa ggacgcaccc gacatgcgcg cctgttacgc gcggccgcac gccaccgtcg 14400
 tgtgctcgcc ggcggccacc ggcaaggct gtccgaattt cgtgcgtcg gattggcca 14460
 tgggacggcc gttgtctctcg acgcgcgaga acggcatcgc ggagatcctc gatcgatcg 14520
 ggcgcggcat catcgcccg cgcgtgcgc cgtcgtcg cgcgcgcgt gacacgctcc 14580
 ggcgcggctg ggccggctgc gccgacgcgc cgcgcgggt cgcagaggag acgttcgacc 14640
 tcggccgctt cgcgcgcgcg tacgacgcgc tgcgtgtt ggtggatcc gccggcgtcg 14700
 gcgagcacgc gaggcgccgc tcttgcgaa gggactcgcc ggacggctgc tcgggtggc 14760
 tcagaaggcg aacgatttcg cgcactacgt cgcacatcga tcgcacatcg agcgcacgc 14820
 gcgggcgaat cgcgcaggc tccatcgccg gtcggccgc cgcgtgaacc ggttcctcat 14880
 cgtgcggctg gccagcatcg ggcacgtcg cgcgcgcac ggcgtcctcg gggccctcg 14940
 cgcgcgtac cgcaggcgac gatcgactt cctgcgtcg gacgcacccg tggccgtcat 15000
 cggcaccac cccgcgtatca acaccgtcta cacgtccag gatcgatcg gcctgtccga 15060
 gtacgactgg atcatcaacc tccagaacct cgtgcgcgcg cgcgcgttc tcagaacgtc 15120
 cgggcgtcg taccgcgaga ttctcgagca gatcgatcg cgcacatcc tctactcgat 15180
 cacccggcgat cacctcgaga acggctcgga gtccacgcgc acgacatcc tctactcgat 15240
 ctgtgagatc gaggagcatt tccatcgccg gtcgtcccg tacgacacgg cccgcgtaccc 15300
 ggtgacccgcg attcacgtgg acgcacaggc acgagcggcg gacgcacatcg agttcgccgt 15360
 cccggagaac cgcgcgtgc tggcgatctt cctcggtcc aactcggtcg ggtgcggccg 15420
 cgcgtggcgc ttgcgtcgatc actccatcgat ctcgtgaaac cgcctcgatc gtcacttcga 15480
 agatcggttt gcgatcgcca tcatcgccgatcc gtcggccatcc ggacgtgatc gatcgatcg 15540
 acagtaccgcg gagatcgatcg gccgcacatcc ggacgtgatc gatcgatcg acaagacgtc 15600
 gctcgacgag ctgcgtggccc tcatcgactc gtttcctcg ctcatcgat tgcgtcgatc 15660
 ccccgatcgatccat ctggcgatgg cccgcggatc gccgcgtatc ggcgtgtacg tcaacgcacgg 15720

gacgttcagg atgcacccga cgctcgccga cgagcgcttc gtggccatca acagcacgcc 15780
gccgtgcttc acctactcgt ggcgatggaa gttctctgc tcgacgtgcc gcgacccggc 15840
gaccagatcg cattactgtc acaacgaggc gttcggttt ggcgtggacc gcattccgct 15900
ggccgcaatc gaccgcgcgg cggcgccgtt gctcaagggg agctccgatg caggacctca 15960
ttcgcgcgca gtttgaagag tcgatcgaga cgaagcgcaa agcgctctcg ctcgtcgatg 16020
cggtggccga gtcggcgccgc ctcctcatcg aggcgctcg ggcggacac aagatcctcg 16080
tgtgcggcaa cggcgctcg gccgcccact cgacgcattt cgccgcccag ctcgtgggcc 16140
gttcgagaa ggagcgccgc gcgttgcctg ccgtggcgct cacgaccgac acgtcgatcc 16200
tcacccgcgtg gccgaacgac tacgcgtacg agaccgtgtc agccggcagg tgagtgcctc 16260
ggagcgccgg gcgacgtgtc catcgccatc acgacgtccg gcaactcgcc caacatcctc 16320
gccgccatgg acgaggcgga aaagaaaacgc atgaaaacgaa tcgcgctcac cgacgcggc 16380
ggcggcaaga tgaagagtct cgccggcgct tgccaggtca tcgtcccgctc agatcagacg 16440
tcgaggatcc aggaagtaca cattaccgtc atccacatct gggcgaagtt gatcgagaaa 16500
gcgcctcagct gagctggcg tagacgtcg acgtcgatc ttcgaacgtc gaccaggagt 16560
gctcgccggc gatgaggtcg gcggcgccgt cggtgaggcg gcggcgagg tgctgatcct 16620
cgaacaacag ttgcagtcg ttcgcgttagc ctcggcgctc gtccgcctgg acgagaaaagc 16680
cggtcttccc gtctccacg acttcacgcg cgcggggat ggtggagccg atcacgggca 16740
cgccggccga ctgggctcg agcaccgaga gcggcagccc ctcgtcggtt gacggcagca 16800
ggaaggcgctc ggccggctgg agcagcaccg ggcgtcctg gcggaaagccg agaaagcgca 16860
cgcgatcctc gagcgcgagc tcggcagaga ggcgcctcag cgcctcgatc tttcccttc 16920
cctcggtccg tgcgtcccc accagccagc actcgaccgg caggccgcgc gacatcagca 16980
gcctggtcgc cggattgcc gtagcctgcc ctttgtgtgg cgcaagattc gcgagcatga 17040
ggacgacgaa gcggtcggtc tcgaggccgg cgccgtacg cgcggcttc cgggtcgccgg 17100
ccacgaactt ctcccgatcg accgcgttcg gaatcgcggt gaccggcacc ttgtccggc 17160
gtccctgtcg cgccgcgtcg atcatcacgg cgatgttagcg cgcacacgtg acgtgtgt 17220
gcggcgggct cttcaggacc caggcgatct ctcgggatc cggtcgatc tgaaagtgc 17280
cgcccacgcg cgcgcgcgcg gcctgcagcg cgggtctcat gaatccgtac acgacaggat 17340
tgtgcacgtg gaccaggggc cggccacga gaccaatcc cgaaacatcg cgcgcacacg 17400
cgccgagctg ttgcaggccc cggcccgca gcgcgttcag gtcatacata cggccgcgg 17460
cgccctccgc gtcgagcgcg gcccgcag gaccctcgcc tgtaatccat gccgtgcacg 17520
ggacgcgcg cgtttggcg gggccgaca accggatcgc cagcaccgca gcgcgcgg 17580
ccagatgtga gacgaggatc tcgtcgaggc gggggactg attcatgtt ggcgcgtatc 17640
cgcccgccc gccagcctac gttactcggtc tcgcaggatcc gcccgcgg tagagcgagc 17700
gcagcaggct gcaagtc gacgggagga agcacacggc gtaccgcgtg agatcggtgc 17760
gcgcgtcg caggccgccc tgaccgaacc gttccgcgc cgcgtaaac tcgcccgggt 17820
cgtactcgag gtggccgagc gcgagcagga gacgatggcg atgatcgatc gcggcgatc 17880
gcccgcactg atcgctcccc cgcgcgcagga accgcctccag cgtcctgacc atgccttcga 17940
gcatcacgt gacgacttc gacatgttgg tgtcgatcg cctgtatgc gtcagggtt 18000
ccggatgaa accgaccggg ccctcgccg agaggcgacg ccagaggatc cagtccctcg 18060
cgccgtggat gcccgtcg aacatctctt gcccggacac ggcggctttg accatgaccg 18120
acgaagcggt cacggcggtt cgcaggacga ggtcgggag acagcgcccc tcgagcggcg 18180
gaagatcgag cgtcgccagg acctcgccgg ctcgtcgat gacggcgccg cccgtatgaa 18240
tgaagctcca gtcggcgatc tgctggaaaga gatcgacctg cggcgaatc ttggccggat 18300
cccacgagtc gtcggcgatc agcagcgcca cgaattcccc tcgcgcgcgc gcgatggcg 18360
gattcgggc gcccggcaagg ccccggttgg cctgagagat caccgtatgc gcgcgcgg 18420
atctcgccag gatgtctgcg tgccgtcgat tcgatccgtc gtcgacgacg atcacttcga 18480
cggcgtata ggtctggag agcgccgact gcacggccct ggcgtatgtag cgctcgccat 18540
tgaacaccgg gacgatcacc gacacgaggc cggggatatc ggtctggcg gccggcgagg 18600

gctctcgagc ggcttcagcg gccatggccc tctgactgg cgtcgcgagc gaacagcatg 18660
acgtttctcg aaagccacgg agacaacggc ggcgttca tccgctcccg cagcgtctgc 18720
gatgccgggc cgaggtagct gaacctcggt gccgcaga tcttctgcca gtactcgagc 18780
ggctgttcgt ttagtggcc cattccgccc tggcccggt ggcggcgct gaagagcacc 18840
agcggcgccg tgttcgtgag cacggcgacg agttctcgca aagcggcggt gccaagggtgt 18900
tcggcgacct cgaggctcat ggagagggtcc gttcggtcccg cggccgtgga caccggctg 18960
agatcgaaatg agttcagcgc gacgcccggc tttccgcccgt cgtaccgtcg cgctgtactcg 19020
gatcgctcg agccggccgc cttcacgcct cgtcgagga acgcgtgcac gaagtgaccg 19080
gtgccgcagc cgaagtcgtt catcgatcg atgccgggtc gcaacgcccag gaccccatca 19140
acgagcacgg ggcggccctc ggcggcgacg tgcccgaaca gatccatgac ggccgcgttg 19200
tacgtccatc gctcgaagcc cagccgttcg ccgatccggg acatcttcag tccggcctgg 19260
caacgccccg accagtagag ccgcttgcgc tgggtcacga tcatgggttc gtcccggccg 19320
cgacgcgaac ggcggccggc gcatgcagcg acggcgagaa ctccacatcg gccacgccc 19380
cccacgatcg atgttccctt accgcccacg tcgcaagctg ccgcgcacggc acgaggaacg 19440
tctgggtcgg gttgtgaaac acgtgcaccc acaggttagt gtcggccgcg acgagatgcg 19500
cgccggaaatg gaggtcgagc gagaactcgc ccatgaacgc ggacgcgtcc ccgagatccc 19560
gccgggtggat gttccgtcg tagacgacga gctgatcggt ggacgcgtgc accatgaagc 19620
cgaacgtgac gtcgtccacg ggctcggtac ttctcgccgt gacctggagg ttcagctgt 19680
tccccggcgt cactcgctg tgcgtcccg acgactcgac gccgacgaaag agcgagccgg 19740
acacgatctc gacggggctc ccgacactcg gctgttccga tttccctggac ggccggacgt 19800
acgcgtcgag gacttcgatg gccgtccgg cggccctcac cgagccggcc agatacacccg 19860
catgctcgca gaggtcgccg accgcctgca ggttgcgcg aacgaggacg atcgcgtatgc 19920
cctggcgctt gaactcctgc atgcgatcga agacttctg ctggAACgtc atgtcgccga 19980
cgctcaacac ctcgtcgatg atcagcacgt cggccgcag gtcgcggccg atcgagaagc 20040
cgagccgcgc gttcatgccc gacgagtagc gcttgcacgg cgtgtccacg aagtcggta 20100
cgccggcaaa ctcgacgatc tcgtcgaacc gctcgccgt ctcgacccgt ttcatcccc 20160
tgatcgccacc ctggaggaac acgttctcgc gcccggtag gtcggatgg aacccggccg 20220
cgacttcgat gagcgcgcggc ggcgcgcgc ggacgatcga gccgcgggtc gtcggcttca 20280
ggatcttcgt cagcacctt ggggtggcgtc acttgcggcc cccgttgcg ccgtatgtgc 20340
cgagccctc gccgcgggtgg acttcgaacg agacgtttct gaccggccag aactcctgtt 20400
cgtccagcgc gtcggccggaa cgcgggttgcg acagccgcg cacgagcgc ggcacgaggt 20460
cccgcaggct gtcgtacgc tcgcccccc ggaacttctt ggagacgttg tcgaagacga 20520
cgggggccgt cacgtcagac gttctccgcg aattcgaact cggagcgtatg gaagaacacg 20580
cagccgaaga cgacgcgtac gaccgagacg atcgcgtga cgcggccat cacggacggc 20640
ggctgggtca ggacgacgc ggcgcggaaag ccgtcgatga tcggcgatc cgggttgcg 20700
tgcagggccc actggagctt gccgcacc acctcgaccc ggtagaccgt cgacgcgcgc 20760
aacatccaga cggcgatgac gacctcgaaag agttatttca cgtcgccgtaaagaggttc 20820
gccatcgaca ggatcagggc gatggccgtg gtatcgtatca cgacgcacggc gacgcacggc 20880
ggcagccaga ggatgttcgg tccgaccggc acgtggtagt agatcatgag cgacgcacgc 20940
accacggagc cgatcgcaaa gtccacggcg ctcacgatca cggccggagaa cgggaagatc 21000
tccctggaa agtacacccctt cgagacgaga ttctgttcc cggtcaggaa ggtcacggag 21060
aagcggagcg ccgaggagaa gaagttccag acccacaggc cgcagaacgc gacacccgga 21120
tacgcgaccg gctgtcgat ggccgcgcg cgcgtatcgaa tcacggagaa gacagccgtg 21180
ttgtatgagcg gcatgaacac ggcggccacc aagcccatca cgcactgtt gtagcgcacgc 21240
agcaggtcgc gccgcgtcat ctgcacgcg acgtcgatc actcgactg ctcggccgc 21300
atctcacggc aatcgccag gaaggcgccg atcgtcagcg gttcgatcg cggcgatcg 21360
ccggccggat catgagggca ggttagcgtcc gagctgcggc aagagcaact gcacgaactc 21420
ggtcgcgcgc cgcgtcgcc gtcgtcgatc gacggcacgg ggctcgatc 21480

gcggacgagc gccggcgtccg atcggttcag ccgcacatcgcg tcgtacacca ttagatctt 21540
gctccagtac tocgctggcga tcacgcggcc gtggctctgg taccagtaca gcacgattt 21600
ccgctcgatc cccttctgga tgacgtatcg gttgatctcg atgggctcgg tgcggcccg 21660
caccttgacc agcgcgcggc cggtggatac gggctccag ccggcgcccg gcaggcagtt 21720
catcgccgag tggatctgtg cgccttcgcg ctggctctgg tagtagccg ttagagcga 21780
cacgtagggc tcgcgcggc ccaggtatcg ccggttatcg tactctgtcca cgccgagcac 21840
cgccatgacg tccttggta agggcgcgct gtcgcgtccg gtccaccgct ccacatctggaa 21900
cgggacgttc gcgagcggt ggcgaagcgg gacctgtcg ggcgcacatcg agcgcgcggc 21960
aaacgcgcgc gtgcccggaga aacacacgct gagaatgtatc agtctgtctgg tcatgaatct 22020
ggactcacac cgtggccggc gcggtcaccg ctggggcgag aggctgtaaag cgccgcagga 22080
cccagaccac gacgaacaac atgacgaacg cgcaagaaa gacgagccag cccgagaacg 22140
tgtggaaagaa cccctgcgcc gcccctcg cgtagtagtg cgccgcacg cccgtgcgg 22200
ccacacgcgc gccgttcgcg ataatcgca tcgggatcgat cgagagcgc atcgccaggc 22260
gcatgcccac gcgccggatcc gtgaagtagc cgtacacgat gcccagcgt agcagcggaga 22320
tgagcgaccg gatcccgctg cacgcctcg ccacttcgag cgctgtgtgc gcgagcacga 22380
tcacgttgcc ttgcgcgcg accgggatgc tgacgcggc cagcacgccc tcgcccacgc 22440
gagacgcccag gagctgcacg ggaaatgcga tctgattgaa gatgtatcgac gggatcggg 22500
tcatcagcac caggaacgcg agcggaaacg cgagcaccca gagcttgcgc cagccgagca 22560
ggaacacgat cgccgggta agcgaggcta gtatcgacac ggcgcgtcaga aagagctcg 22620
cgccgagcaa ccccgcgacg agcagcgcg gcccggcgag cacgtatcgac agaccgagcg 22680
cgctccggcgt gtccggcagc gcccggcgac gcccggcgctt cccaggca aagtagggcg 22740
cgagcggta gatcagaatc ccgtggagt agttgtcgatc gttgatccag tcctgcgcca 22800
accgcacgat cacctgcaca taaaggagca ggaaactcactc ggcggcgata ccgagcggaa 22860
ccagtttgcgctt ctgaagcaca ctcatagcga tcgggacgaa tatacggaaag ggctctccc 22920
ggcagaactt gggcgccgaa acacgttaacc atcacgctca caggaagttt ggcggtaaat 22980
gacgggattt gccggcttgg gggcctcgatc gaatctaccc aatggatgac aaatatccct 23040
atcaactcct cttttgact gcgtacgcg gaagatcatc ggcctttca gatgcaggac 23100
aaatgtcatc tgcaggaaatgatc ctggccacag cgccggtcaa gtggaaatccc 23160
ccgacagaat cgcgtaccac cgcaacttggc atccggcgag aaaagtcgc tcgtcgcc 23220
gaaaactgca ttacaggaaac gccaggctgg cagccgggtt gtttccgtc agtattttt 23280
ggcattgcag ggcgtttcgatc ctggccgggg gcgaaaccgt ggcggccggc cgatcagcgc 23340
tggagcaggc cagggacagg catggatggt tccgtgagac agccgcgttc gggagacggc 23400
cgggctcgcc gggagctcgatc ccacctggcc gtggaccgcg cggcgccgcca atttcctgac 23460
gcggccggccg taagtttga cgggtctcg ctctcgatc gcccgttgc tcggccgc 23520
aacgcgtcg cgcaccgcct gcaacgtcat ggcgtcgatc ctgacgtggt ggtcgccatc 23580
tccctcgacc ggtccatcgatc gtcatcgatc gggatctcg ccgtgttcaaa ggccggccggc 23640
gcctacgcgc cgctcgatcc tgcgtatccg aaggaccgc tcgagttatc ggtcaggac 23700
gcccggccgc ggatcgatc ctggccgttgc gcaatctcat cggctggcag 23760
gtcgacgtca tctgcctcgatc tgccggacgg cccggggcgatc tggcgaaaga gggccatgcg 23820
ccggacgcgc gcgatcgatcc cgaccaccc gctgtacgtca tctacaccc gggatcgacc 23880
gggcggccgcg agggcgatcgatc gatggccatc cgtccgttgg cgaatctcat cggctggcag 23940
gtcgacgtca cggccgtcgatc cggccggccgcg cgcacgttcc agttcacttc gccgagcttc 24000
gacgtctcgatc tccaggaaat ctttgcgacg ctcgtacccg gccggcgatc ggtcgatcgatc 24060
tccgaggaca cgcgcgggatc tccgcgcgcg ctgcgtcgatc tgctgcggc tcgatccgtc 24120
gcccggctgt atctcccgatc cgtcgatcgatc cagcagcttgc cctgtacccg caccgacgat 24180
ccggatctgc cggcgatcgatc tgaggatcgatc acggccggcg agcagctccg ctgcacgc 24240
gcgtcgatcc agtggatcgatc gccggatccg cagtcacgc tccacaacca ctacggcccg 24300
ggcggacgcgc atgtcgatcgatc tgcgtacacg cttgtactaccc ccccgcgatc atggccggcc 24360

agaacaact ccaccgggg ttcgagctcg gaccgttgcc gccgctcgcc ggcgaggctcg 30180
cggcgtcgtaaaggattc acggggatgg agcgcgttgg attctccaac accggctccg 30240
aagccgtgct ggcggccacg cgaaaaatggccaaacccgtcac cgggcgcac aagatcgccg 30300
tcttcgaagg gtcgtatcac ggtctcccttgc acgaagtgc cagccggccg ctcgtggta 30360
acggcgaatt gcggctggcc cccgcggcgc cggcatcgc cggcagcgcg gtgagcgaag 30420
tcatcctgct cgagtacgcg aacccgaagt cgctcgaggt catccgcgcg cacgagtcgg 30480
agatcgccgc gatcctcgca gagccgtgc agagccggcg gtcgacactg cagccgaagg 30540
agttcctgca cgagggttcgc cgcgtggcg acgagatcg cgcgcgcgtg atcttcgacg 30600
aagtcatcac cggcttccgc gtgcaccccg gggcgccca ggcgcacttc ggcgtcccg 30660
cgatcttgc cgcctacggc aagggtcgacg gggcgccat gcccattcgcg atcgtcgccg 30720
gcaatgcgaa gttcatggac gcgttcgacg gggccactg ggagtaacggc gacgggtcg 30780
tccccggagggt gggggtcacg tactttgcgc gcacgttcat gggcatccg ctcgcgtcg 30840
ccgctgcgaa ggccgtgctg acgtacatga agtcgcaggg tcccggcctg cagacgcggc 30900
tggcagatcg gatcgcggc ctcgcgaacg acgcgcgcgc gatcgtggcc aggacacggc 30960
cgccgtatca gattacgcag ttcaagtcga ttagtgcgc gggcatccg ctcgcgtcg 31020
agacggcggt gtcgtgttc ttccgtatgc gcgagcgcgg gattcatatc tgggaggggcc 31080
gtcccttctt cttcacgacg gtcacacccg aggccacta cgcgcgcgtc ctcgcgtcg 31140
tcgacgagtc gtcggccgag atgcaggcgg ctggcttctt cggcgccccc gcccactccg 31200
cattgcaggta gacccggacg ggggaccagg acgtcctgcg gttcaccgaa ggacagcagg 31260
agatctggct cggcgccag atggcgagg cggcatcacg cgcctacaac gaagtcgtcg 31320
cgctcgaccc ggcggggccct ctgaatcgcg cggcgatgca ggcgcggcgtc gacgggtcg 31380
tcgcgcgtca cgaatcgctg cgcacccg tggccagcga accgcgcggc ctcgcgtacg 31440
ttccgggac gtcgggtccg gttgaatggc aggacgtgtc gggattgcgc gaggacgc 31500
agcacgagc cgtgcagaag gtcgtcgagg gggcgacgg cgtggcggtt gacgtggaa 31560
ggggggccgtt gtcgcgtgt accaccatga cgcgcgtcc cggcatcac gtcgtcg 31620
tggcgccca ccatctcgta tggacggct ggtcgttcgg cgtcatccgc cgcgcgtcg 31680
cgacgttcta ttccggccag gttcgaggca cgcgtgcgaa cctcgaggcg cggatgcagg 31740
tgagccgctt cgcgtgtac gatcgcggc ccaagcagag ttcgtaaacggc gccgagaccg 31800
aggcggtctg gatccggcaa ttgcacccg tcccccggcc gtcgagttt cccggccgatc 31860
ggcccgccgccc gcaatgcgaa tcgtaccagg gggccgtgt gtcgggtccg tttgacgcgg 31920
cgctcgctcg cggcgccaaag aagtcgcgg ccaacacccg cacgaccctg tttacgcacgc 31980
tgctcggtcgat ttccgagacg ctcgtctacc gactgaccgg ccagcaggat ttcgtcg 32040
gcaccccggtt cgcgtgtacg cctctcgatc gcgaggatct gtcgcgcac tgcgtcaact 32100
tcgtgcgtctt cgcgcgtccg gttcaacccca cgcgcgtac gtcggagttac ctgcacacgc 32160
tgaggacgc gtcactcgac ctcaatgcg accagaactt cacgtacggc agttgtcg 32220
ccaagctgcg gttggcgaaag caccggggcc gcaatccgcgt cgtctcgacg agttcacgc 32280
tcgagccgc ctcgtcgat cccgggttcg acggcttcac ggcgcgttc ctcgcgttc 32340
cgcggtcgac atcgaagcgc gacccgtacg tgaacgtcat ggagatggac ggcggctcg 32400
tggtcgaggc cgcgtacgc accgatctgt tcgtacggacc gacgggtcgcc cgggtggctgg 32460
atagctatcg gatcctcgatc gaggccgtcg tggatcccg cggccggcc ctcgtcg 32520
tgccgggtat ctcggagccg gaccggagtc agttgtcgac cggctggaaac gacacggccg 32580
ccgactaccc ggcgcgtccg gtcgtccatc aactcgatgcg aagacgcgcg gtgcagaccc 32640
cgcggtcgatc tggatcccg agccggctac gtcggatccg ctcgtcg 32700
gcaatgcgaa gttggcgcat ggcgcgtcg aagacgcgcg agcagggtgg 32760
cgctgtcgatc ggaacgtacg gccgggttc ctcgtcgatc gtcgtcgccg cacaaggccg 32820
gcaatgcgaa tggatcccg gaccggatcc ctcgtcgatc ttcgtcg 32880
aggattccgg tggatcccg gtcgtcgaccg attcgcggccg cgaaggccg tggatccgc 32940
ccaggacgc ctcgtccac ctcgcacccg acaggccgcg aatcgcgcgaa gatcgcacgc 33000

atatctgaa ggtcgccctg tcggccgaag acggccgt a cgtatctac acgtccggat 33060
cgacgggaca accgaaggc gtcgtcatcg ggcacccca gctcgtaat tacacgtgg 33120
cggtgatcg a cggatcggg ttacccgccc gtcgagtt a tgcgctcg t cgaacgtcg 33180
cgccgatag cggtgacc gggctctcg c gtcgacgt gggctgggtg ctgcacgtga 33240
tacacctaga ggtcgacc gacggggggg cgctcggtc gtatggacc gttcacgaca 33300
tcgacttcct gaagatcaccg cctcgctc tcgcgtcgct gtcggcgat caccctcg 33360
cctcgctc gccgcgcgc tgtctcatgt tggcgccga ggcgtacgt cccggcggtgg 33420
tggaaagact gca gctc gcaccaggct gcccggatcat gaaccactac gccccgacgg 33480
agaccacggt cggcggttg acgtactgga cgggtgaccc caggatctg ttcgtcgaga 33540
gcgtgcggc cggcaggccg ctcgggaacg tccgcgctta cgtgctcgat cccggccggag 33600
cgcccgcgc gattggcgtc gtcggcgacg tctgtatcg cggcgctgc gtggcacgacg 33660
gttatctcaa tcgtcccgac ctcaccggc a g c g t t c g t c c c c g a t c c g t t c g c c a c c g 33720
agccgggagc ggcacatgtac cggaccggcg accggccccg gttccgggccc gacgggaaca 33780
tcgagttcct cggtcgcgcg gactatcagg taaaagtacg cggattccgc gtcgagctcg 33840
gcgaaatcga a g c g g c g c t t c g c g c t a t g a g g c g t c g a g c a g g c g t a g t c g t c 33900
gaaaagatca gcccggcgc g a g c g c t g a t c g g c t a t g a c a g c c g g c g c t 33960
cgatctcgat ggcggagctt cgcacgtcgc tcaagcagag gttccggac tacatgggtc 34020
cggcggtat cgtcgctc gacaagctgc cggtcacgtc gattggcaag atcgatcgac 34080
gcgcctgcc ggcgcgcgc g a g c g c g c a g c g c t a t g a c a g c c g g c g c t 34140
ccgagacgga ggcgcgcgc g c g g a g a t c t g a g c g c g g t g c g g t c t c g a g a a g a t t g 34200
gcgcgcacga caacttcttc gacctggcg ggcactcgct gtcgcggcc cgcaccgc 34260
tgaaggctcg cgacacgttt cgcgtggacg tgcgtctcat cacgttcttc gaaatgccc 34320
ccatcgccat gatggcgca gcaatcgacg cggattccgc ggcgcgcgc tcccgacca 34380
tcacgcgcag ggcgcgcgc a g g g c c c g c a g g a t g a c t g a g c t g a g c t g a g c 34440
ccacccgcag actgagtccc cgcacctgcg ggcgcacggg atcgccgcgc a a g c g g g c g a 34500
cgtgtttctg ggcgcgcgt cgtacccca gggcgcttg tggccctgg cgcagctcg 34560
gcctgaaagc tccgcctaca acattccgat cgcgtcgat c t g a a g g g c c t g a c c c 34620
cgcgcgcgc g a a c g c g c c a t c a a c a c c a t c g t c a g c g t c a c g a g a c g c t g c g a c g a c 34680
gttctcgctg cagggcgatg agctcaagca ggtcatctct cgcacgtcgc gtattccgat 34740
ctccctcgat gacatgcgtc tgctcagttcc tgagaaacgc gaggccggat ggcgcgcgt 34800
tgcggccgc gacggcgccg cgcgtttcgat tctggcttag ggcgcgtgc tcaaggctcg 34860
actcgatcatc ctgcagccct cgcatttcgt ctcctgatc accgtgcatac atatcgat 34920
cgacgggtgg tccggcgaaa tcttcatccg c g a g c t g g c c g a g g c t g a g c t g a g c t 34980
cgacgggcgc gacccgaagc tgccggagct gccgatccag tacggact ttggcggtt 35040
ccagaaggaa cagaccgaag ggcgcgcct gccggctgt tgaaagact ggacgcagga 35100
gctggcgaaa g c g t c g a t g a c g t c a c c t g c g c a c g g a t c g t c g c g t c a 35160
gaccccccgg ggcgcgcgc ctcgtttcg gtcgacggc gcaatcgatc cggcgctgc 35220
cgcatttcgc cggggcgaga acgcgacggt gttcatgacg cttctctccg cttcgccgc 35280
cgtgtcgat cgcgtgtccg gacaggaaga catcctcatc ggcacccgt cggccggccg 35340
cgatcgatcg gagcttgagg gctgatcgg gttcttcatac aacacgttc tccgtcggt 35400
caacgcgtcg ggagatccca cgttcgcga gttcgatcgg cggcgccgc ggcgtgcgg 35460
tgcggccatc ggcgcacgg acctggctt c g a t c g g c t c g t c g a g a c g c t c a a c c c g g a 35520
gcggacccgg gatcgatcgatc c g a t c t a t c a a g t c a t g t c g c g a g g a t c c g c g t c 35580
gaagatgacc ttggccggaa tgcgtatcgat g c g g c t t g t c g t c a c a a c c a a c c a a c c a a 35640
gtgcgacccctc acgtcgagg ttggccgaaa c g c c g a c g g c g t c a c c c t g t a c g t c a g g a t a 35700
cagcgatcgatc ctgttcgtatcg cggcgacggt ggcgtcgatc ctgcagcaat atcgccgtgt 35760
gcttcagacc g c g a t c g c g g a c c c g g a c c a g c g g c t a t c c g a g t c g c t c g t c a g g t c g c g a 35820
cgacgacgcg atgcagatcgatc tgcactggg gaccggatccg g c g g c g t c t a c c c g c g c g a 35880

cgtcagtctg gccacggtct tcgaagagcg ggccgcgcg actccggcgg cagtggccgc 35940
gacgctggag ggacacgacc tcacgtacgc ggatctcaac cggcaggcga accgcctggc 36000
ccggcggttg aaggcggtcg gtgtggggcc gaacgtgctg gtcggcgtgg cgctcgaccg 36060
gtcggtcgat ctgctggtgg cgctcctcgc cgtggcgaag gcccggcgcg cgtatctccc 36120
gctcgatttc gtcacccgc aggaacgct tgcgttcatg ctcgccgaca cgcgcgcgc 36180
catcggtctc acgctgcggc gatttgcgg ggcgtcgag tcctttcccc tcttctgcct 36240
ggacgacgag ctgcccggca cggccgagga gcccggacaa gatctgcgc cgcagtcgac 36300
cggagaccac ctggcctatg ttagtgcac ctcggcgtcg actggtcagc ccaagggcgt 36360
cgcgatcacf catcgccgcg tgggtcgct cgtacgcggc acagactacg tccgctggaa 36420
cgacgtgcgg gcgatcctcc acatcgccgc gacgtcggtc gacgcgttga cgttcgaggt 36480
gtggggcgcg ttgctgaacg ggcgcgcgt ggtgctggtt cccgagcggc tcgtcagcct 36540
ggagacgctc gagcgcacgc tccggtcgga gcaggtggac tgcctgctgc tcacgaccgc 36600
gctcttcaac gccgtcateg acgcgaagcc cggcattctc gtcggcgtga agcaactct 36660
gatcgccggc gaagcgctgt cagtcgcgc cgtgcggcgc gcccggcgc acctgcccga 36720
cacgcggctc gtgaacgcgt acggtccgac cgagtgcacg accatcgctg ggcctaccca 36780
gattccgcgc acgctcgatc cgcaggcgcg gtccatttcg atcggccggc cgattgccaa 36840
tacgcaggcg ctcattctcg atcgacacgg cggcattctc cgttcggca tcgcccggca 36900
gctctgcctc ggcggagacg ggctggcgcg ggaataacctg aatcagccgg cgctcactgc 36960
ggaacggttc gtgcggcaacg cggtcagccc cggccggcc gcccggctgt accgcacggg 37020
agacctcgcg cggatgcgc cgcacggcaa catcgagttc cttggccgga tggatgcgca 37080
gatcaaactg cgcggcttcc ggatcgagcc gggcgagatc gagcatgccc ttcgtgcggc 37140
gggagacgtt ctggacgcgg tgggtggagat cccggcgtat gcccacatgg cccatctcg 37200
cgccctacgtc gtgcgcgcgc atggagatca actcaccggc accgacactgc gcgaacgtct 37260
gaaggcgcgc ctgcggaaat actgtgtgcc cgcgaagtac gtgttcctgg cggaggtccc 37320
gcgaacggcg gcccggcaagg tcgatcgctc ggcgtgcct gcatcgccga tcgaggcgc 37380
cgcccccggag cggcgcgatc cgtcgttcct gaacgaggtc gaagcggagc tcgcccgaat 37440
ctggagcgaa gtgcgttcggc tggccctccgt gctcgcgacc gacaatttt tcgatcggt 37500

<210> 2
<211> 37507
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: DNA sequence
of clone FS3-135.

<400> 2
gatcccacga tcggcgcgcag cggattgcgc agctcggtcg cgagcatcgc gaggaactcg 60
gtcacgtggc ggcgcctcgcc ctccagctcc tccatcgctc tgcggcgtcg gagggtcgcc 120
gtcaccttgg cgaagccat gtgccttcca ctcgcgtcgc gaagcgccgt cacgacgacg 180
tcggcccgaga agcgcgtgcc gtccttgcgc acgcgcgcgc cctgttcctc ggcgcggcgg 240
tgcagcagcg catcggtcgag ctgcgcctcc gggcgccgg ccgcggcatc ctcggccgg 300
tagaagagcg aaaagtgcgg tccgatgatc tcttcggcgc tggtagccgt gatgcgtcg 360
gcgccttgcgtt tccagctggcgtt cacgtggccc gcggggtcca gcatgttagat cgcgtgtcc 420
tgcacgcctt cgtacaggag ggcgcgacgc tcctcgctt gcccgcgtcg ctccctcg 480
tcgcggcgct cggtcagggtc ggcgcgtgacc ttgcccgaagc cgaccagcgt ggcgtccctcg 540

tcgcgcagcg ccgtgatgat gacgttggcc cagaagcgcc ggccgtcctt ggcacgcgc 600
cagccctcg cctcgaagcg gccgtcctcg cgtgcgcga gcagtcctt ggccggccag 660
tcgcgcgca tgtcctcctc ggtgtagaac gcggagaaat ggccggccat gatcttcg 720
gcccgcatacg cttcaactg ttccgcgc gcggtccacg aacgcacgta gccctcgaa 780
tcgagcacga agatcgctgta gtcgtgacg gacccgacca gcagccgaa gatggcgtcc 840
tcgcggcca gcccgtcgg ggcgcgaag gcgaggttcg gcacccgaccc 900
gatggccttgcatggctat taggttacaa cggtgcacgc aagttcaat cggacaatca 960
gagataattt cttgtatgaag atgcatttt tggattttt acgcaggag ctggggacgc 1020
tcgaggccca gggccgtac aagtccgagc gcggtcgtca ctctccgcag ggccgcagcg 1080
tgcagggtggc cgcccacgac gtgctcaact tctgcgcctt taactaccc ggcctggca 1140
accaccgcgc cttggtcgc gccgcgaagg agacgctcgac gcgcgcacggc tacggcatgg 1200
cgtcggttcg cttcatctgc gggacgcacg cggcgtccacg cgagctggag acgcgcctcg 1260
ggcttgcgac gggcaccgag gacgcgttcc ttacgggtc ctgcttcgac gccaacggcg 1320
gacgttgcgac gacgtgtcg ggcgaggagg acgcgggtat ctggacgcg ctgaaccacg 1380
cctcgatcat cggcggcgtg cgcctctcgaa aggcaagcg cttccgcgtc gacaacaacg 1440
acatggcgag cctcgaggcc cagctgaagg cggcggacgc ggccggcgcc cgctacaaga 1500
tgatcgccac cgacggcgtg ttctcgatgg acggcatcgat cgcgaacctg ggcgcctcg 1560
gcgagctggc cggcgctac cacgcgtatgg tggatggta cgactcgac gccgtgggt 1620
tcatggcgcc gacacggccgg ggaacggcccg agcattgcgg tgcgttgcgg aagggtggaca 1680
tcctcaccgg cacgtcgcc aaggccctgg gcccgcgttc cggcggctac acggcgccgg 1740
agcgcgaggt ggtggcttgc ttctcgcaacc gctcgccccc ctatcttc tccaacacgc 1800
tgatgcccgc catcgccggc gctgcgtca aggtgtcgat tctcctcgag ggccggcgcc 1860
agctgcgcgc gaagctcgcg cgcaacgcgc gccacttccg cggcggatg acgcgcctcg 1920
gttcacgct ggccggccgca gaccatccga tcatcccggt gatgtggcc gaggccccgc 1980
tcgcgaagga gatggcgac cggctgtga aggaaggcat ctacgtggc ggcttctcg 2040
ttcccggtt gccccgaggg caggcgcga tccgcacgc gatgtggcc gcccattgac 2100
cgaagcacgt cgatcgccatc atcgccgcct ttgccaaatg cggccgcgtat ttgggagtca 2160
ttgttgcgaa accctttcgaa agacgaagcg cggccggc atctggatgg tcgagtcgac 2220
caagccctgt gtcggccaca acgacgtgt gatccgcgtg aagaagacgc ccatctgcgg 2280
caccgacatg cacatcttc acgtggacga ctggcgtcag aagaccatcc cggccgtat 2340
gacggcggc caccgtatcg taggcattgt ggaggccatg ggccaggagg tgcgcggcc 2400
gcaggcgtgc cagcgcgtct cccggaggg ccacatcgatc tgcgcgtt gcccacactg 2460
ccgcgcggcc cccggccacc tgcgcgcctt caccgcggc gtggcgtat accggcccg 2520
cgcttcgcg gactacctgg tgattccgc ggagaacgcg ttcccgattt cgcacgcac 2580
ccccgacgag atcgccctcg tccctcgatcc gttcgccaaac gcccgcaca cccgcgttc 2640
gttcgacgtg gtggcgagg acgtgtgtat caccggccgg gggccgtat gcatcatggc 2700
cgccgcgcgc ctcgcacga agatgggtgc cggccgcgcg gtggacgtat cgaaggaaaa 2820
cctgaaggac gtgtgcgcg agctggcat ggtcgaggcc ttcgacgtcg gcatggagat 2880
gtcggcgtg ccctcgccct tccgcagat gctgcacacc atgaaccacg gcccgcac 2940
cgccatgcgtc ggcacccgc ccaggaggg cggcgtcgatc tggacccagg tggatctca 3000
gggcctggc atcaaggccg tccacggccg cggatgttc gagacgtgtt acaagatgtat 3060
cgccatgcgtc cagactggcc tccacgttccgc gcccacgtt acgcacccgtc tgcacgtcg 3120
cgactacctg aaggccgttgc agacgtgggg ctccggaaaa tccggaaaaag tgcgttgcgtc 3180
gtgggattag cgaggccac ggtcgccgc cgggttcccg cggccgcgtat cgtgcaggcg 3240
gttggcagc acgccccggg atttccgcgt tggctcaatcgt ggcacttcgc cgtgcacgc 3300
ccgatcgccga tcgacgtcgat cgtgacgtcg tgcgtgtcg atgcccaggac cacctcgatc 3360
tggacgcgt cgtggccgg caggccgcgc gacacccgt ggctaccgc cggcacgtat 3420

cgccgcctccc tgccgttccgc gccgttccgc ttccggccg cggccgcctc ggtcgagttc 3480
 ggcgtctcga tgcgggtgca aggtgagcgc cgcgtctgg ccacccgcgc gggcggtgt 3540
 cccgcggcg caccacccacgg cgtggcgccgc gcccgcctggc acctcgagac cctggatggc 3600
 acgcggcgcc tcgagacgc ggcgttggcc gatccacgc acagctgggtt ctccacgc 3660
 ttcgaccatcg cgtcgccac gtgcgtcgag tacctggcg cccgatggcc cgggtggcg 3720
 gggcggtgc tggatgcggg atgcggcgac ggcacacgg cgcgtggcat cgcgtcgcc 3780
 tatgcgcggc agcaggtgtt gggcatcgat cggggcgct getaccgggt gcttcccgac 3840
 atcctcgagc gccacgcct cggccaccc ggcgtgcgg acaacctgca gttccctccc 3900
 gcgatgcac acgcgttccc gttcgccac agcagcttcg acgtcatcgat ctcgtggcc 3960
 gcggtcgagc atttcgttggg cggctaccc ggcgttggcc cccgaggcg cgggtgtcc 4020
 aagcccgccg gcctgtgtt catccatccc gagotctact acaccgcga ccacggccat 4080
 cacctcgccg agtacagccg cggcccttc ttccacctgg tgaagtgcgc cgacgagggt 4140
 cgcgcacatcg tggcgccgc ggacgttgc ctcacacgc cggcgccat cgcgcgcac 4200
 cgcgcggagc actggcgctg gtacaacgc ctcacatcgca tcacggccgc gcaatcgag 4260
 gacgagggtgc ggcgtctcgat cttcgagccg tggcgctgg cactgcgcgc cggccgt 4320
 gtggagtaca agccggagct gtcgttgc cgttgcgtac agctggcgat ctcggagctc 4380
 tacgtggctt gcatcaaccc caagccgtg tcaatgggtt cctacaggac gcatggcg 4440
 gccgcattcc ggttaagggtcc gcccggacgg gaattgaagg cgcacgcgc ttgttatccc 4500
 cgtcacttcc ggactgcgg gatcaccctc acaggagaac ccccgatgaa gacccagacc 4560
 caggccgtgc tggcgccgc cctgtgtcc cttgcccggca ccacgctggc cccgcggaa 4620
 gccgtcacgc tgacggcgcc gagcgaatcc cggccgtca ctcgagccgc gaaggcgac 4680
 ggcacggtgg tcgtgaaggc cgattgcacg gtcacccgcg agatcaccgt caccggcatg 4740
 accgcacccg cggccacat ccacgaaggc aaggccggc ccaacgcggg cgtggcagtc 4800
 ccgttgcgtg agacccgcgac cacacgttc gaggcagccc cggcgccaa gatgaccgag 4860
 ggcgttgcg ccgttgcgtgaa ggccggccg acttacgttgc acgttgcacag cgaggcccac 4920
 aaggccggcg aggtccgcg ccaacgttgc ggcgttgcgtt gcaatcgatc caaaaagaaa 4980
 aagcccgccg aagccggctt ttcgttgaa ggctttgggtc gggcgaaag gattcgaaacc 5040
 ttcgacccccc tgacccccc gtcgggtgc taccaggctg cgttgcaccc cggccggatc 5100
 tgaaattata cctgcacatcg ggtcgaccc gccatgttgc gaaacgcggg gggcttcggg 5160
 tagcatccgc ccatggcgac gaccctcgat acgggtgggtt caggctacat cggccacccac 5220
 atcctcgatcg cactggccca ggcggccgcg cgcacatct gcatcgacaa ctactccaaac 5280
 agctcgccgc gttccatcgac acgcgttgc cagatcgac cccgttgcgtt cggcccttc 5340
 gacgtggaca tccgcgacgc cgcacggatc cgcacagggtt tgcgtggcc cgcgtggac 5400
 agcgtatcc acctcgccgg actgaaggcg gtcggcgaat cggtcgagca gcccggccgc 5460
 taccacgaca acaacgtccg cggcaccggag acctgttgc cggcgctcg cgcactcgccg 5520
 gtgcgttgcgtt tccctcgccg accgttgcgtt gtcgttgcgtt gacccgttgcgtt 5580
 atcgacgagg acgcggccac ctcggccgcg agccgttgcgtt gccaacaa gtcgttgcgtt 5640
 gaggacatgc tggcgccct cggcaacgcgc gacccttgcgtt ggcgttgcgtt gacccgttgcgtt 5700
 tacttcaacc cggtcggccgc acacgagac ggcgttgcgtt ggcgttgcgtt gacccgttgcgtt 5760
 cccacaacc tcatggccat cgtgttgcgtt gtcgttgcgtt ggcgttgcgtt gacccgttgcgtt 5820
 gtctacggca ggcgttgcgtt caccggac ggcaccggca tggcgatccatccacatcg 5880
 tgcgttgcgtt ccgaaggccg cgtggccgcg ctggcgacac tggcgccgc gggccgggg 5940
 accgttgcgtt cgggttgcgtt cgggttgcgtt cgggttgcgtt cgggttgcgtt 6000
 acctcgacgc ggcgttgcgtt actaaggatt gcccggccgc tggcgatccatccacatcg 6060
 gacgttgcgtt tccctcgccg accgttgcgtt gtcgttgcgtt ggcgttgcgtt gacccgttgcgtt 6120
 aagcccgccgaa tggcgatccatcgaaatccatcgac cttggcgccgc gggccgggg 6180
 ggctacgggtt gacccgttgcgtt aatccatcgac cttggcgccgc tggcgatccatcg 6240
 agggggcggt tccatcgatccatcgac gggatgcgtc caggacgtca aaaacacgcac 6300

ccggcggtgg aaggaagaag gcaaccctag tgggtcatg gctccagacg ggattgcgcg 6360
ccgagtaaga atcgtagtc gagaccaact cgcctccggg tgcaacgaat cgtgagccga 6420
agcaacaccc cgcacatcggt cgccgtcgcc ggcattgcgt tccgccttcc cggcgctacg 6480
cgcaaatccc tctggccctc gtcctcgag gggcgcaacc tgcgtacccac cgtcgatccc 6540
gcccgtggg cgaaggagcc gtactaccac cgcgcgaatg cgcagccgg caccagctac 6600
tcgttcgccc cggcaccat cggcgacatc agcgcttcg acgcgtcggtt cttcgccatc 6660
tccccgcgcg aggccgcga gatggatccg cagcagcgcc tgctgtcgta gctctccctgg 6720
gaagccctcg aggccgcggc ggtgaagccc tcgaccctgc gccggcagcga ttgcggcgtg 6780
ttcggtggcc tgcaccgt cgactacagc tacggctcg cctacgaccc ggcgtccatg 6840
gacgcgtcg cggccaccgg caacaccggc agcategcg ccaaccgcct ctcttatttc 6900
tacgacccctc acggcccccag catggtcgtg gatacggcgt gctcgctctc gctggggcg 6960
ttccaccagg cctgccatc gatcgccatc ggcgagacgc gccaggcgct ggtggcgcc 7020
atcagccctc acctgcattc gtacgggttc atcgcttct ccaaggcgtc gatgctctcg 7080
aagcgccggc cctggccgt gttcgaccc acggcgatg gctacgtcg ctcggaaaggc 7140
gccccgtgc tgggtgtgaa ggacctcgac cgcgcgtgg cgcacggcaa cccgatccctc 7200
gccccgtcg cgcgcggc cgtgaataacc gccccccgct cctccagccct caccgtgccc 7260
agcgtcgatg cccaggctgc attgctgtcg gatgtgtacg agcgcggccg catcgaccccg 7320
tccgagatcg actacctcgaa ggcccacggc accggcaccgg tgggtggcgag tccgatcgag 7380
acccgcgccc tcggcatcgc gctcgccgcg ctgcgcggcgc gcccacagcc gtcgcgatc 7440
ggttcggta agagcaacgt cggccacccctc gaggcgccct cgggcattggc gggcctggg 7500
aaggcgctgc attgcctcaa gcaccgcgcg gtgcgcggca ccatccaccc ctgcgcggcc 7560
aaccggcaaca tccacttcga cgagtggaaac ctgaagccgg tccctcgagcc gtcacgctc 7620
catccgtaca agaagctcg cgtcgccgtg aactcggtcg gttcgccgg cgccaaacgcg 7680
cacgtgatcc tcgagagcac gcccacttcg cccaaaggcg tcgaggcccc gcccggcgag 7740
gtggcgctca tgcgtggcgag cgagcaaccc ggcgtgcacg acatcgctga caccaccgtg 7800
gcgaacttcc tgcaggcgag cgagcaaccc ggcgtgcacg acatcgctga caccaccgtg 7860
caactgcgcg actggcacca gctccgcgc atgggtgtcg gcacccggc cggccagcatc 7920
gccccggccc tgcgcattt cgcgcggcc cacaacacgc cggcgctgt gaggcccg 7980
gcccgcggca acgcgtccgg cccgcattt gtgtattccg gcaacggcgc gcaatggccg 8040
gcaatggggc gcccctgtc cgcgcggag ccgggtttcc gtgcgcggct gcaatcggtc 8100
gacgcgtgt tccgcggat cggcgctac tccatccctcg agatgtggaa gcaggaggc 8160
cccgcgcacc agctggaccg caccgagatc gcccaggccg tgcgttcgc cgtgcagggt 8220
ggcttgaccg agctgtcgcc ccaactgggc atccagccca cgcgcgtcac cggccacacg 8280
gtggcgagg tgcgcgcgc ttggcgagc ggcgcgtga cgcgcgtac tgcgcgcgc 8340
gtgatctacc accgcagccg ctaccaggggc gagaccaagg ggcggggcgc gatgaccggc 8400
gtcgccctcg cggaggacgc cgcgcgcgc cagatcgagg cgctcggtt gaaggccgag 8460
gtgaacgtcg cgtgcattaa cagcacgcgc aacgtgcacg tgcgcgcac ggcgcgcggc 8520
atcgaggccg tgcaggccga gtcacgcacg cgcaagggtt tccaccggcc ctcgcaccc 8580
gactacgcgt tccacagccc ggcgtggac cgggtgcgc acgggtcg gacgcgcgtg 8640
cgccgcctca cgcgcgcgc caccgggtc ggcgtccatt cggcggtcac cggctcgccc 8700
gccgcgggca accagctga cgcgcacccatc tgggtggagca acatccgcga gcccggcg 8760
ttccaggccg cgcgcgcgc catcgccgg tgcgcgtca atgtcttcat cggatcgcc 8820
ccgcaccggc tccatcaagaa ctacatcaac gacggctcg ggcgcgcctc gatcgaggcc 8880
cgccgcgtgg tcacgtcgac ggcgtcgccg agcgatcgac cggcgatccg cgcgcgcgc 8940
caggaagtgc tgcgtccgg ctggccgggt ggcgtccat cgcgcgcgc agatgtccc cgcgcaggc 9000
cacttcgtcg agctggccgc ctaccggc tgcgcgcgc gccactggcg cgcgcgcacc 9060
tcgcaggccct acgacccat ccacgcacggc aacgcacacc cgcgtcggtt ctatccctc 9120
cacgagaacg acttccactg ggaaaaccac atcgacacccg cgcgtatcc cgcgtacgc 9180

gaccatgtcg tcggccgcgc cgtggtgttc cccgcggccg gcttcgtcga gatggcgctc 9240
gccgcctcg cgatcgcgt gggccggcgc ggcacgaga tcgagacgct cgagatccgc 9300
agcgcgtgc tgctcgagga ctccacgtcg aagaccgtgc gcttcgcgt ggaacccgt 9360
ggccgcttca caatccgcag cccgcgtcgc ctgagcgagg atccgtggca actccacgt 9420
gtcggcaagc tggtcggcac gcccaccggag ctccgcgcg cgcgcgcac gtcggccgc 9480
tcgcgcaagc ccgatgtgat cgcggcccg cactacgaag ggcgcgcgaa ggccggcctc 9540
gcctacggcc cggccttcca gtcggtctcc aagggtgtgc tcgacggccg ggacccgggc 9600
agcgcctacg cgcgcctgat cctcccaag ccgattcgcg gcgagctggg cgtgatgcac 9660
ctgcaccccg cgtcgctgga cggctgcttc cagctgctgg tcgacctgct ggcgcgcgag 9720
gcctccggc acgcgcagat cgcctcggt ccgatctgg tcgggcgcac ggcgcctac 9780
ggccctgcgg ccctcgtaac cgcctgcgc gtcggcctga ccgcgcgcag cccgcgcctc 9840
ctggtcgcgg acttcgagct ctacggccccc aacggcgacg tggtgccgc gtcaccggc 9900
gtgcgcgttcc gcccgcgttca gctgaagcag ccgcaggccg cgcgcctcg ctagctgtcg 9960
taccgcgcac tcccgccct gaatggccgc gacacgcgg ggcgcgcgc gctgcgcctc 10020
gcgttgcgttcc cccgcgcgttcc gggcgagccg tggcattcgcc ttcggccgg ctcgcgcgaag 10080
cgctactacg atgaggtcgaa gcccgcgttcc gacgtgcgtt gcaagcttcc cgcgcgcgc 10140
gcccgcgcgcg aacggatcgcc cggaggcccg ctcatcgacc cggccgcgc gcccgcgtc 10200
acgcggccgcg agcattggcc gtcctcaag cgcgcgcgc agatgcgtt cggaggaccag 10260
ttgctcgagc cccgcgaagg cggctggccg tggagccgg tcaccgaatt gcccgcgcgc 10320
caggagagct ggatcactcg gtcgcgcac taccggacg aggccggcca gctgatcg 10380
ctcgccgcgcg cccgcgcgttca ctcgcggaa gtgttcgcgc gcaaggaaagc gtcctgcgc 10440
agcctgttgc cggccacgc ggcgaacgtt ttcgcgcgc tttgcgcgcg cgcacccgc 10500
ttcgcccaacg cccggcgccg catcgccgac acgctggccg tcgcctcgac ggcgcgcgc 10560
gcccgcgcgcg cgcgcgcgtt gtcgcgcgtt acgcccgttcc gtcgcgcgcg cgcgcgcgc 10620
gtcctcgagg cccgttgcgact cggccgcgc gacttccgcg ttcgcgcgcg cgcgcgcgc 10680
gcactggccg agtacgaagg ctcctcgat gccattccgc acgctgcgc ttgcgcgcgt 10740
gacctgaaga tcccgccat gggcctgaag gcccgcgc acc agggccgcgtt cgtatgcgt 10800
atcggttcgc atggcctgct cgggtggccg gatccggatg cggccgcgc gcacccgc 10860
ggcgcgcgcg cggaaagacgg gtcgcgcgcg atggtcgcgc agcatccgc ggcgcgcgc 10920
gaccccttc tccgcgcgcg cccggcggtt tggacgcgttcc ggcgcgcgc gtcgcgcgc 10980
tcgcgcgcgc gttcgcgcgc cggatggccg atcgcgcgttcc cccacccgcg cttcgccgc 11040
accgtgcgcgcg tgcccgagct gcccggccgc gaggccgcgc gtcgcgcgc gtcgcgcgc 11100
cgccgcgcgcg cgggtggccgaa ggctgcgcgc cgcgcgcgc ggcgcgcgc gggctggatc 11160
ttcgtgcagg aagcggccgcg ttattcggccg gcttcggccg ctcgcgcgc ggcggatct 11220
cgccgcgcgcg gccacgcgcgtt gttcgagatg ggcgcgcgc gtcgcgcgc ggcgcgcgc 11280
gcccgcgcgcg cgcgcgcgcgc ccagacgcgtt ggcgcacatca gcccgcgcgc cttcgcgcgc 11340
ggcctgcgcgcg acgacgcgcgc gcccacggccg gacgcgcgcgc gacgcgcgcgc gcccgcgcgc 11400
gcccgcgcgcg cccgcgcgcgc gcccgcgcgc gcccgcgcgc gacgcgcgcgc gcccgcgcgc 11460
gcccgcgcgcg cccgcgcgcgc gcccgcgcgc gcccgcgcgc gacgcgcgcgc gcccgcgcgc 11520
gcccgcgcgcg cccgcgcgcgc gcccgcgcgc gcccgcgcgc gacgcgcgcgc gcccgcgcgc 11580
gcccgcgcgcg cccgcgcgcgc gcccgcgcgc gcccgcgcgc gacgcgcgcgc gcccgcgcgc 11640
ctcgagcgca acgcccgcgc gtcggcgcgtt gacgtgcgttcc cccgcgcgc gacgcgcgc 11700
gtggtcctca cccgcgcgcgc cccgcgcgcgc gtcgcgcgc gtcgcgcgc gcccgcgcgc 11760
tccgtgcgcgc gcccgcgcgc gcccgcgcgc gcccgcgcgc gtcgcgcgc gcccgcgcgc 11820
aagaaccccg cgtggccgcgg cgcgcgcgc gcccgcgcgc gcccgcgcgc gtcgcgcgc 11880
gagggtgcgcgc cccgcgcgcgc cccgcgcgcgc gcccgcgcgc gtcgcgcgc gcccgcgcgc 11940
gacgaggccgcg tccggccgcgg ctcgcgcgc gcccgcgcgc gtcgcgcgc gcccgcgcgc 12000
gtcaccgcgcg tcggcaagga cgtgaccccg gtcgcgcgc gcccgcgcgc gtcgcgcgc 12060

gccccctcg cgttcgccac gcacgtggc accaccgccc attcggtcgc gaagaagccc 12120
gccccgtgga cgttcgagtc cgccggccacc atcccgaccg cgttcttcac cgtgtactac 12180
tcgttgaagc acctggccca gctgcgag ggcgagcgcg tcctgatcca cggccgcgcg 12240
ggccggcgtgg gcatcgccgc gatccaggc gccaagtggc tggcgccgga gatcttcgac 12300
accggccggct ccgacgagaa gcgcgactc gtgcgcctgc tggcgccgga ccacgtgctc 12360
gactcgccga cgctcacgtt cgccgacgac gtgcgcgca tcaccaacgg cgagggcgtc 12420
gacgtggtgc tgaactcgct ctccggcag gccattgcgc gcaacctgcg cgcgtgcgc 12480
ccgttcggcc gcttcatcga gctcggaag cgcgactact acgagaacac gcacatcgac 12540
ctgcgcccgt tccgcaacaa catctcgat tcggcgctcg atgcccacca gotcatgaag 12600
gagcggcccg acctcgccgc ccagctttc aacgagctga tgcagctt cggcaggcgc 12660
gtgctctcgc cgctgcctca ccgcgccttc gcccaccg aggccgtgg agccttcgc 12720
tacatgcagc actcgccca gatcggaag gtggtgcgt cgttcgccga cggcgtgaag 12780
gcccggcccg cccggccgcg cgtgcgcgc gagctggcgc tgccggcga cggccacgtac 12840
ctcgaccccg gggcttcggc ctgcgcaccg cgcaactggc ggtggacaag 12900
ggtgcggcc acctggtgct ggtctcaag agcggcgcg agtcgtcga gaccaaggcc 12960
gccccggcc acctcgaggc ccaggccgc acgggtatcg cggaggcgtc cggccatcacc 13020
gaccgcgcct ccgtgcaacg gtcgtcgcc gaggtcgccg ccgcgtgcgc gccgtgcgc 13080
ggcgtgatcc acgcggccgc cgtgatccag gacggcttca tcaccaacat gacggccgc 13140
cagatccgcg acgtgtggc gccccaggc ctcggccccc gccaccttga cggactcacc 13200
cgccggccca agctcgattt cttcgtgt tactcgccg cgacgacgct gttcgcaac 13260
ccggccagg ccaactacat cgccggcaac tgcttcctcg aggactcgc caaggcgcgc 13320
cgcgccagg gccagccggc gctgtcggt ggctggggcg cgatcggggc cgtggctac 13380
ctcgccgc acgagaaggta gaaggaagcc ctgcagacgc acatggccgg caccgcactc 13440
gagtcggagg cccgcgtcgc ggtgtggag cagctcctgc tgccgacgc ctcggccctc 13500
ggcgtgtcg acttcagctg ggcacaccc cggccgttcc tccccaccgc cgcctcgcc 13560
cgcttcgcg agctggcgc ggcggcgac gacgcgaagg acgacgacga cggccgcgc 13620
gagctgcgc gcctcgccaa cgagctggac gcctccgagc tctcggtat gttcaccgac 13680
ctgtcgcc gcgagggtgg cgagatccctg cgcattccgc cggaccgcct cgatacgcgc 13740
cagcccttgc aggaaatggg catggattcg ctgtatggcg tggagctgt gaccgcgtg 13800
gaggcgcgtc tccggcgtgaa cctgcccgtg atgtcgctt ccgagcagcc gtcgatcgag 13860
aagctggtc accgcacatcg ggcgcgtcg aaggacccga acgcggcac cgaggccaa 13920
tcgcgcacgc accagatcg ggcgtcgcc ggcgcgtacg cggccgcgt cgactcgcc 13980
caggcgtcagg agctcacgc ggcggcgtat gacgcgcgaag ccacgcagca gggacgtccg 14040
tgaaggcgcg caaccccgcc ggccttcgcg aggcaatcgaa ggacgcgtg atccagcgc 14100
tgctcgagca ggcgtcgcc cgggtcgaga aggacaagcc tccggccgtc gaggcccca 14160
tcaacccctt ccggccgcac cccgatgccc tggggccacat ctggagcgc tactgcgc 14220
tcgacatgca cccggcttac cagcagatgc agctggtgaa ggaagggtcc ggcggactgg 14280
gcattcccgta tccgtacttc cgcgcgcacg acggcgtggc cgggatacc acgcgcattt 14340
ccggccgcga gtacatcaac ttctccagct acaactacat cggcttcgc ggcgaccgg 14400
cggcgtatgc cggccgcac caggccatcg acgaaatcg gaccctcggt ctcggccagc 14460
aggctcgatc cggcgtcgac ggcgtcgac cgcgcgtgg agcgcgcgt cggccaggc 14520
tatggcgtgc acgacgcgcgt ggccttcatc acgcggccacg cgaccaacgt ctccaccatc 14580
ggccacctgt tcggccgcgc cggccatcg tgcacgcacg cgttcgtcga caacacatc 14640
ctgtatggcga tccagcttc gggggccaa cgcgtggcgt tccggccacaa cggactggcg 14700
gccctggacg agctgctcg cgcgcgcgc cggccactcg agcgcgtgt gatcgatcg 14760
gagggcgtct acagcatgga cggggactac cccgagctgc cggattctc ggcctcgcc 14820
cgccggccacc ggcgttcct gatggtgat gacgcgcact cttcgccgt gatggcccg 14880
cgccggccacc ggcgttcct gatggtgat gacgcgcact cttcgccgt gatggcccg 14940
cgccggccacc gcatccgcga ccacttcggc atggccggcg acgaggccga catctggatg 14940

ggcacgtct ccaagacgct cgcctcgta ggcggttca tcgcccggcga gaaggcgctg 15000
gtcagacacc tgaagttcgcc cgccccccggc ttccctctaca gcgtcgccat ggccccggcc 15060
gtggcggccg cggcgctcgcc cgcgctgaac cgcatgcgcg aggagccggc ggcgtggcc 15120
gcctcgagg cgccggacg cttcttcctc gaggccggc ggcggcgagg catcgacgtg 15180
gcctctcgcc agggcatggc cgtggtgcgc gccatcaccg gcagctcgat cgcgcccgg 15240
cgctggccg aggcgatgtt ccagcgccg atcaacgttcc agccgatcgat ctaccggcc 15300
gtcaggaga actcgccgcg ctcgcgttc ttctgtggcg ccacgcacac cgaggagcag 15360
ctgcgttca cggtccggcga gtcgcgcac gcctggcgca agctctgagt ggcgggcccc 15420
cggtgcgcg tcgtcgctg caccacggc ggcctgaacg ggcgtgttgc gctctcgccg 15480
ctgctggccg cccacgcgt cgaagtctcc ggcgtggta tctccagccg cgccggcggt 15540
gccatgagt cgtatggccg tgccgcgtc ggatacgtgc ggcgcggcgg cgtgcctat 15600
ggcgtgtacc tctggtgcgc cacggcgctc gccgacctgc tgctgcgcgg cacgtccgag 15660
ggcccggtcg cgccatcgcc ctcgcgcgc ggcatcccgc tcctcgccac ggcgcgcgtg 15720
aacgatgcca cccgcgcgc cttcatcgcc gggccggcgc cggacctgat cgttcccg 15780
ttttcaacc agcacatcgaa cccgcgcgtg gctgcgtcg cgcgcgtggc cgcgttcaac 15840
atccaccgt cccgcgttcc gcaactccgc ggcgtggatc cgggttccctt cgcgcgcctg 15900
cgccgtgcgcg agccacacgg cgtgagcgatc catcgatcg aacccggctt cgataccggc 15960
gcgtgtcgcc cccaggaaac cgacgtcgag gccccggca ggcgtttgc cgcgcgcgcg 16020
gcgtctatgc accggggggc ggcgttccgc gccggccgtg cgcgcgcgtt ggcggccgac 16080
ccgcgcggaa ccccccagcc ggcggggggc ttctacgact cctggccac cgcgcgcctg 16140
gtcgccgcct tccggcgccg cggccggccg ctgcgtcccg cccgcgcac gtcgcgcctc 16200
gcgcgcggg ggcggccgc ttctgttaata gaatcagcgc ggtagccgg actcacgcgc 16260
cgctttcccg cggccggccg aggccatgaa aacactggct gaagcaacac gcaatttcg 16320
cgctcaccgt ctcgtgcgc accgtggccg cgatcctcta ttccgcgcctg atcgccctcg 16380
acgtctacat ctccgaatca cgcgttgcgg tgaggagccc ccagcgccag gtgcagaccc 16440
gcctggtggg cgcctgttc tcgggcaccc gcttcgtcg ctccaggac gacacctact 16500
cggtgcacga ttcatcacc tcgcgcacg cgctggccga gctggacaag aagctcgccg 16560
tcgcgaagct ctacacggcc gccaacatcg acttcataa cgcgttcccg ggcgtcgact 16620
gggacgacag cttcgaggcg ttccaccgt actaccagaa gaaggtcact atcgacttcg 16680
acaccgcgtc ctgcgtacg gtgcgtcgcc tgccgcgcctt cgagaaggcc gactcgccg 16740
gaatcaacga ctcgtgtcg cagatggccg agccgtggta gaacgagctg aacgagccca 16800
gccgcgcagg cctgatccgc ttccgcagg ccgaggcttc gctcgccgag gacaaggta 16860
aggatgccgc gtcgcgttc tccgcctcc gcaacca gtcgggttcc gagccgcacc 16920
gccaggccctc gatccagctg caggccgtgg ccaagctcca ggaggagctc atcgccaccc 16980
aggggcagct ggcacagctg cgcaagctt cgcgcacaa cccgcacatc ggcgcgtcg 17040
agaacaagtc ggcggcgctg cgcgtggcga tggcgccgca gtcggccaaag gtgcacccgg 17100
gcagccggctc gttcagcgcc cgccggccgg cgttcgtcgcc ctcaccctc gagaagggt 17160
tcgcgcaccc ccagctggc gttccctca ccgcgttggc gaccgcgcgc acgcaggccg 17220
agaggaagca gcttacaccg gagcgcatcg tgcagccaa ctcgcgcac gaaatcggtcg 17280
agccgcgcgc catccgtcg atcttccaccg tggcggtcg gggcttggc gtcggccgc 17340
tggtgagcct gtcggccgtg acggccctgt ggaccgcggc cggcatgaac cacggctcg 17400
cttccctctc ggcgcgcctc ggggtgcagc ggcgcgtcg ctacgcgtcg ctgcgtcg 17460
aggatcatcac ggcgttccgg cgcgcacacc tcggcggtcg gtggcttggc gtcggccgc 17520
tgatcttac gtcggccgtg acggccctgt ggaccgcggc cggcatgaac cacggctcg 17580
cgctgcgtat cgtggccctc gcggtcaccg gatactccgc ggtgtggta tggcgcaact 17640
gcgcctcgcc cgcctcgatg gccatcgagg ccaacaaggg ctcgcgttc caccgcac 17700
tgcgcgtgtat cgacgtgttc gtgacgcgc tcatcctcgat gatgcggccg ggcacggcgt 17760
cgttcaccgt gtcggccatc ttcttcatct ggatcggtcg gatgcgcgtat cgcac 17820

tgctcaaggc cgccgttcggc tggttcatgc tgcgtgggtt cggcccccgc ctggcgctcg 17880
ccatcgccgc cggcaccgcg tactcgctgg tggtcgagcg gctgtggccg ccgacggcc 17940
acctcgctgtt ccccgctctcg ggccgcgcct tcatggtcga gtggctgcgc gagaattcc 18000
agcagttcgt gctttgtctg ccgatggtgc acggcaccga gctgtgcgc gacggctact 18060
tcggcaacgc ggtgcgcacg cactacgacg tggctacat ggcaacggtc tcgctggcc 18120
tgctcgctgt cggcatgcac ctggcgca tggcgccaa ggcgtggag gccgcgttat 18180
ccgggtcgag aacgtgacca gatgtactca cgcgcctcg cccagaacgt gctgacgcgt 18240
cagcttcgag ctggccgcg ggcgaacat cggcatcctc ggcaagaacg ggcgcggcaa 18300
gtccacgctg atccgcctga tcagcgccg cagagcccc accaccggc gcatcgctcg 18360
cgagatgagc gtgtcctgac cgctcgccctt cggccgcgc ttccagaccc acctcaccgg 18420
gctggacaac ctgaagttcg tctggccat ctacggcgctc gactaccgcg acaaggctcg 18480
ttcgtcgag gacttcaccc agctcgccgt gtatccgcg gagccgggtt tccactactc 18540
gcacggcatg atcacgcgc tcgccttcgc gctgtcgatg ggcgtggagt tcgactgtt 18600
cctgtcgac gaggcgatgg tggtggcgaa cgcgcgcctt cagcgcgcgt gccacgtcg 18660
gctctccac aagcgaagg accgcgcctt catcctcgac acccacgacg ccaagggtat 18720
caagctctat tgcgagagcg cctgcgtgt gcacgagggg cggctgctgc cttccccac 18780
cgtggacgccc gcctacgagt tctacatgaa cgaggtcatg caggacctcg ccccgaggt 18840
tgcctgaacg ccccgccgc ggcgcgcgc gtgcctcg agctcgccg ggcgttcgac 18900
ggccacgccc gcatcccgca ggaggcacgc ctgcgttcc gcccgcgtcg catgtcgag 18960
ggcatcgacg tggagggct gtcgcgcac agcggccacg tgctcgccaa gggcctgcgc 19020
ccgcgcggcg gcgccgcacct cgcacccggac cgcgcgcgtga accgcctgtc ggcgcgtcg 19080
gtctccacca agcaggagct cacaacgcg cactgtgcga cggcgcgtat ggcattcgc 19140
cagctcggtt ggcgcgcgc gaccctctac cgctcgacg cggcgcactt cgcgcacttc 19200
gtctggcagg cgctcttcgc ggcacgcgtc cgcgcgcgg acttcgactc gtcacccgc 19260
gcggcttc gcgtcgccgc cgtccgtgg acggccatgc accgcgtcg gtcgtcactg 19320
cgcaagctcg gctacacgct gcatccgcgc atcgacactt cgcatttcga cgtgcgtatc 19380
gcgcgcgc cctaccggc ggggtcgcc gacggcacgc gcctcggtt ggcgtaccac 19440
gacgcgtacc cgcctctat gccgcacacc atctccgaca tgcctacca ccaggcctcg 19500
caactaccggc cgctcgccgc caacgtggcc tcggcgccgc acttcgcgtg cgtctccgaa 19560
gccacgcgc accgcgtct ctcggcttt cccgaggctcg aggccgcgtc cagcaccatc 19620
cacaacatgg tctcgacca ctacttccgc gagacgcgcg acgcggggcg catcgaggag 19680
atccctcgaa cgcgcgcgc cgcgcgcac aaggcgccgc ccaacggagc cgcaccgc 19740
gtcgacgat cgcctccctt ctcgcgcgc gccgcgcac accgcgcaccc cggctacctg 19800
ctcgccgtgg cgacgatcg gccgcgcga aaccacgcgg gactgtatcg cgcgtggag 19860
cacctcgca cctcgcgctt cccgggcctg cgcctgggtt tggtcgccat gcccggctgg 19920
ggctacgagg cgctggtagc caagttcaag cctgtggctcg cgcgcggcc gctcttcgt 19980
ctcgaggacg tgcccgcccc ctagctgcgc ctgcgtact cgcacgcgc tgccacggtg 20040
tgcccgagct tcggcgaggg cttcgatttc tccgcgtgg aggccatgcg ctgcggcagc 20100
ccggtgatcg cctcgagat cgcgcgcac cgcgcgggtt accgcgcacgc cgcgcgtac 20160
tgcaaggccctt atccctggc cgcctcgcc gagccgatcg ggcgcgtcat cgcaccggcg 20220
gcgcacggcc tgcgcgcaggc gctggtcacg cgcggcaccg aggtctcgca ggcgtacacg 20280
cccgaaagcca tcctcgccca gtggcgccag tacctgtcg gcacgggtcg cgcggaggcc 20340
ccgtgagcgat cccgaaggcc accgaccagg cgtgggagga gtggggccag cgcgcaccct 20400
acttcggcgat gatcacaac cgcgcgttcc gccgcggca gatggacgag gaagcgaagc 20460
gcgaattccct cgcgtccggc cgcgtccatg cgcactacgt gatgcgcacat gtcacgcgc 20520
acatcgccgc cgcgtccgtt cgcgcacca tcctcgattt cggctgcggc gtggggcc 20580
tcgtcatccc cttcgccgc caggccgcgc aggtgaccgg cgcgcacgtc tccccctcg 20640
tgctggccga ggctcgccgc aactgcgcgc agcaggcgat ggcacatgcg cgcctcg 20700

tctcgacga ctcgctcacc gggctgcggg gcccttcga cctcgacac tcgttcatcg 20760
 tggccagca catcgatccg gcgcgcggc gcgcacatctt ccgcgcctg ctgggcaccc 20820
 tggcgagcg gggcgactggc gcgttgcact tcgttatgc gaagcgatc tacgcggcga 20880
 cgtacggcg ggcgcaccc cccgagccgc cgccgcggcc gccgcggcc cagcctccac 20940
 ccagccggc ggagatcaag gccgcgagca aggccggc cgcgtgccc acgcaggcac 21000
 gggcgccgg accggcaccc aatcccgacc ccgacatgca gatgaattcc taccggcgg 21060
 cggagatgct gttccctgt caggaagccg gcgtcacgcg cttccacgtc gagttcaccc 21120
 accacggcg cgagctgggt ctcttcctt tcttccgaa gcccaggccg tccttcgagg 21180
 cgaaggagag gcccagcg cgtatcgctg cggcgctg cggggcgacc gccgcgcgg 21240
 cccgcacttc ggcagccacg cgctcgaggt cgcgcgcga cccgtacagg cgacgcgc 21300
 agtagtgcac ccagtaggag tccggacagg ccgacgcgt tccacgcgc cgctcgac 21360
 cggggccgac ggcgcctgg tgcacttggt gcgaggccca ttgcagggtt acggagggt 21420
 cgagggcag gacggcgacg agcgcctcg tcaccgcag cgcgcgcgc tcatgcgc 21480
 cgtcgccgag ggcgcacgc gtcggcagg actccggcgc aagcgcgtc gccgcgcga 21540
 gcaatcgcg cagcatcgca gcgtgcgcgc gcgcgagcgc caggcggcg cgtggtagt 21600
 gggtccagaa cgaaggcgcc gcccacaaca gtttcgcctc cgcgcgtgc gccggcacgc 21660
 ccccgagag cgggacgcg gcccacaaca gtttcgcctc cgcgcgtgc gacgagaagg 21720
 tccgcaacgg cggcagcacc gcggtggcga cctcgccgca accctcgccg gacgagaagg 21780
 ccgagcgcgc gaaactgcac agcgttgcg cagacgcga atgttccggc cagagccgc 21840
 cgggatccac cgcgcgcgcg gcgcggcgg cacgcgcgcgc cgcctctcg tcaccgcgaa 21900
 gcgcgtcgag cccgcaggcga tggAACgcgg cccactccgc cgacatcgac gccgcggcc 21960
 gtcctcgcg accctcgacg ggcgcacccg tgcgcggcg cgcgtgtgc aggtggcc 22020
 aggtcgccctg cgcgcgggg aggtcgctc cttcgccgac gagcgggtcg aggaccgc 22080
 ggcgcgcgc ggcgtcgccgc gcgcactga agtcgttggc caggccgatc agcgcggcc 22140
 ctagctcgccgg ggcgcgcgg aaggccgcg cgcgtcgccg cccacccgc gcgagggtgc 22200
 tcagtgaatc gaacgagcgc tcgaccccg cgcgcacgtt gccgagggtcc acctgggtcg 22260
 cggccacac ggtcgccgca tgcgcggcca tcggctcgcc gggacgaa ggcgcgttga 22320
 tcagctcgac gtcgcggcaag cgcgcgcgc gcaacgcgcg ctccaggccg ttccgcgc 22380
 cgaggctcgcc tttgtactcg cgataggcat cgcgtcgcc gtcgtattcc tgcgcggac 22440
 tgcgtccat cagcgaggcc gtgaggccgt gtcgcggcg aacgggtcg tatgcgttag 22500
 cgcgcgatc gtcgtaggcc agcgtggccg cgcgcaggcga cccgaacgcg ggcgcgtcg 22560
 ggaatccag cttgcggcgc acgcggggcg ccaggctcg ggcgcgcgc tcgcggcc 22620
 tccagctcgcc cagcgccggc acgttcgagt agagccacaa cggcgtcgcc tagagctcg 22680
 ccaggctcgcc gtcgcggccg ccctcgccct cgaagatcgcc ttccactcc gcgagtccct 22740
 cgtggccggc ggagaagatc ggcacgtgg cgcgcaggat cgcgcacgc tagggccgc 22800
 cgcgcaccc gagacgctcg accagcgac gcagcgcgc gtcgtttcg gccatcaccg 22860
 cgcgtactt gtcgtcgcc ccctcgcc cgtggccggc gagaagggtc ggcgcacgt 22920
 agcgcggctc ggcgccttg tcgcccaggat agggtccgt gccatcagc gagacgcgc 22980
 acacgaagcg cggccgcgaa tcgtcgagct cgcgcacac ggcgcgcgc acggccgaat 23040
 cgcgcgtcgcc tcgaaactgg cgcgcgcgc gggcgaacgc ctgcgcgcg tggAACgcgg 23100
 cgaagccgag cttcttgcac acgcgtcgcc ggtttagaa ccagcgggtgg aacggatgg 23160
 tggccacgcg cgcgttagcc tggcgccgc gcatcgccg caccgcatac acccggtcg 23220
 gcaactgcgc catgttagggc agcgcgcggc agccttcctc gcccaatgcg atgcgggtga 23280
 ggatctcgaa ctcgggtcg cagggttgc cgcgcacac cttcgcgcgc aggccggccgg 23340
 tggcgccgc cggccgcgc gcccgcggc acggcaggc gtcgcgcgc acggcgcacgc 23400
 cgaacggccg caggctcgcc agcgcactcca tcacaggaa gatgtgtcg gggcgtccg 23520
 cggtggccgg cccgcgcgc gaggctccct ggcgcgtccag cagccgcgc acgcctcg 23580

gcagccgccc cgccgcgacc gggcgggcca tgctcccgcc gaccagcacg gctgtcagga 23640
cgaagcccag gcccgcacg ttcacgttgt ggtcggtggg cgaatacgtc atcgcgtcga 23700
actccaccgg ccagcggcgc aagcgtccgg cgacgaaggc gacgcccggc gccggcggc 23760
cgagcagcca ctgcccggca tgcgtgtacg cgtcgtgcag gtcgatggcc tcggtgatct 23820
cgaacgcggc ccacgccagc gccgcggcagca ctgcccggc gcaaggcgcg cgacaccggca 23880
cgccgcggcc atgcaggtac gcccggcgcgca acaccagcgc ctcgcggcgcg ctcaggaaat 23940
cgccccggcc gaccacgcgg cccaagaacg cgacgaggcc tgcgttccacc gctgttaggcg 24000
cgccccggccag cgcacccggc cccaccgtcg cgacgcggcc gaccgcgacg aagagcagcg 24060
catggaaagat ggcgagcagg cccgcgttcc agggcaggcg cccgcgcgcg cccacgggtt 24120
ccacgtcata gcccaccacg agggcggcgt cgacggcggc ctgcagcacc gctgttaggcg 24180
gcacgcagac cagcgtgcgg cccagcagca cgcggaggaa acccgccaaa ccagccgtgt 24240
cgtcgaggat catgggacca ccagggacgt cttccggcca gtctataacg ccccgcccc 24300
tggaaaaacgg cgaggcccgg gggcctcgcc gtggagggaa cccgcgaacc tcagggttgc 24360
gggggtgcga cccgaccagcc gccggggccg atcagggttg cgctgcgtat ggtggcgagc 24420
ccgttcatca ggccgaccgt gacgctgcgg tcgggtgtggc ggaagaccag gtccttgcgg 24480
ctgtcgccgt tcagggtccag cagctgcgtc acgttccaaac ccgtgcggc cggcagcgc 24540
tcgccccggc tggatgtatggaa aacaccatcc atcaggcgga tgcgtacgcg gccgtcggt 24600
tggcggaaaga ccagggtcgtc cttcgcgtcg ccgttccaggat cgcggcaccag gttcaccgac 24660
caccggctgc cggccggcag gagctggcgc ctggcggcga aggtggtgcc gttcatcagg 24720
aacagggtcgc cgcggccgtc ggtgtggcgg aagatcatgt cccgcgtcc gtcggccgtt 24780
aggtcgccca ggtggcttac cgtccagccg ctggcggcgg agaggaagcc cgcgtccggcc 24840
gtcaccgtgg tggccgtccat gatgtatgt tagccggcgc cgtcgctgtg catgaagacc 24900
aggtccgcct tggccgtcgcc gttcatgtcg cccgcctggg tgagcgtcca gccggtcccc 24960
gccccgaaga gctggcgct gccgatgtatc gtctgcgtat ccatcaacca gatgtgcgcg 25020
cgccgtcggt tggcgccag gatgtatgtcc gccttgcgtat cgcgttccag gtcggccgtt 25080
tggctcacga tccagccgag cccggggggc agcagcttct tgccggccgt caccgtatgc 25140
ccgttcatga tggatgtatggaa ggcgcggcca tcgggtgtggc ggaacaggat gtcggcgcgc 25200
aggtcgccat tggagtcggc cgtctggcgtt accgtccagc cccgcggcggc gccgatcagg 25260
ttggccgagc cgggtatggc ggtgcgttcc atcgtccaca cccgcgtatgc cccgtcgct 25320
ttctgcagca cgagggtcgat cctgcgtcg ccgtcaggat ccgagacgc cggcttgggt 25380
ggcggcgatcgat cccacccggaa aaccggcagt gccagccgtt cgcgttccaggc gccgtcgccg 25440
cagttggcgc ctcgggggttc cagggtggcgc agttgtcaagg cgtggagagg 25500
aagttcgaga aacgcacac cagcgtgtgc gcccggcgg tcacccggaa cgcacaccgc 25560
ttccagcccg aggccgtggt gcccgcgtcg ctgcgtatgc cccgtccgtc gatcaggaaac 25620
tcgaacttgc cgggttggg gaagctcgac accgcgtatgg cgaacgcgc gttgcccgcg 25680
agcagcgtgc cccgcaacga gaggtcggttgc ttcaccgtcg tcggatgttggt cgggtcgctc 25740
gtcaccaccc gcccggcgcg aaggctggcgtt gcccgcgttcc accgcgtggc 25800
gtccatcgat tggccggcc cgggtcgat ggtggcgat gtcgtatgc acccgatggc 25860
cacggcgaga ggtatgtatggcc ggtcgccgtt gcccggccgc cgtatgtatggc acccgatggaa 25920
cccgatcgat tggagagcgc gacgttgcgttgc ttcaccgtat accgtcgctc 25980
gccacccggaa cccgtatgtt cttggcgat gtcgtatgc tggcccgat gacgtatggc 26040
gaggtcgccgg tggatgtatggc gcccggcgtt gcccggccgtt tggaggtggc gtagctgacc 26100
gagatggcgc cccggcgcgc gcccgtatgc ctcacggat gtcgtatgc gcccggccgtt 26160
tccggccgcg cgaagggtgtt gcccggcgtt gtcgtatgc cccgtatggc gggcgggggtg 26220
accgtggact tgacgcgcgaa gaggtcgccgtt cgggttgcgttgc ttcacggat gtcgtatgc 26280
gcctgcgcgg tggccgcacgg ttgcggcgca cccacccgg tcgtgcacga gaggcccccgg 26340
ttggagaagc ggttagacccgtt ggtcgatgttgc ttcacggat gtcgtatgc ggtgggtgg 26400
ttccggccgtt cgggtcgat gtcgtatgc cccgtatggc gcccggccgc gcccggccgtt 26460

gagccgttgg gcaggccga gttcaggtg aggtgcgc tcgcgcata gaagtggccg 26520
tacgactacg cgaacgcacc ttgcggcggg ctcgcgtgc cgccgccttc ccaggcctgc 26580
gagggcggt cgtggcggt gcccattggca tggccgagct cgtggatgaa gacccactcg 26640
caaccgaaca ggcagccga agcaacggcg taggcaaagt tgggtccgg cgtgctgg 26700
ccgatccagg ctgcgcccgtt gccgcccgaag tcggcgcgt tgccatgaa ggccaccatg 26760
tccgcgcgtt attgcgtgcg gatggcctcg atgttgcga acgtcgcggc gtcgaagctc 26820
gcgtcgcgcg ggggtatggc atgcattgcgc gtgggtcgt cgatggcg 26880
acctgcgtcg cattcaccag gcgcagcg 26940
ttggcgcgcg tgacgaggaa gttcaggcgc gtcattcaggc tgccgcgcg 27000
aagccgcgcg agtacacgat catcaggcgt atgacgttgcgc 27060
gcgagccgac ggtattcaccgcg 27120
cgacgcgcga ctgcgagccg gtcgtggcca cggcgtggc 27180
gcatgctggc ctgcgcggc gtcatgtcg ccaggtatgc 27240
ggcggtaactc accctgcggg gtcgagaaca cggcgaagga gcttcaggc 27300
cgatcgcgcg gtgggtatgc 27360
cgccgtggc ctggagcaac tcgaagacat acgagtgcgc 27420
gctcgacccgc ggagcgcggc 27480
gcccggcaac ggcgctgccc 27540
ggggaaaccgc ggcgaaggcc acggaaagaga 27600
ctgtatgtcg cggagatgtg gacaaaaggt atcagcgcgg 27720
gggatgtcc aaaaggacta cgaggggtct acggctgg 27780
gaactcctt tggctccgc gcgacttaag gttcgcagg 27840
ggccgatcaa gtccgcgtc 27900
gcccattgtt gtggtggaaaccacgttca 27960
gcgtcaccgt ccagccgcg aacggcggc 28020
cgatcgtcg ggcacgtgc 28080
cgatcgtccgtt gagatctccc accaggttca 28140
cgatcgtccgtt ggcacgtc 28200
ggcggaaagac gaggtcgccc 28260
agccgtgcc 28320
atgcgttagcc ggcacgtc 28380
tgatcgtccgtt ggcacgtc 28440
tgatcgtccgtt ggcacgtc 28500
gtatcgtccgtt ggcacgtc 28560
ggggaaacagc tccttgcgc 28620
cccgtcggtg tgatcgtccgtt 28680
tgatcgtccgtt ggcacgtc 28740
cgatcgtccgtt ggcacgtc 28800
cgatcgtccgtt ggcacgtc 28860
cgagaaaacc cgtgagcaac 28920
acgtcgcggc 28980
cgatcgtccgtt ggcacgtc 29040
ggaacgcgtc 29100
cctgcgtgt 29160
tgatcgtccgtt ggcacgtc 29220
gatcgtccgtt ggcacgtc 29280
tggcaggcc 29340

cggtcacgtt gtcgcccgtg ccggcgagga acgttgaata cgcgaccggg cccgcgctcg 29400
caccgatctt ggtgacgaag ccggccacccg gttgggtctg gtacgcattc acgttgcgga 29460
acgtcgagga gcgcgtggag cccgcgacgt acgtgttgcc cgcattgtcc gttgcgatac 29520
cgccggccgat gtcatcaccg ctgcccgcga gatacgttga atagacgaag cccgcgttga 29580
acgagagctg gtgacgaacg cgtcgcgatt gaggccgggtg ccgacgaaca cggctgcagc 29640
gcgcgtccgc cggtcggaa gccgcgcgcg gccgaacccg tgacatgcat gaagccgttg 29700
ttgtccacgg cgatcgccat gcccctcgat tcgcccggagc cggccagcag cccgcgagaag 29760
atgacgttgc cagtggatt cagcgcgtg acgaatgcat cgcgcgcgc gcatgggtt 29820
gtgcaatcggtt gacggcgaa gcaatgcggg aacgcgttgg tgccgttga accgggtatc 29880
gcgacgaact cgttgcgcgc ggaggccgcg agcgccagcc cggggagata cgcgcgcgc 29940
ccgagcgtgg tcgagttacac cacctgtgc gtgggtgggt tgaacttac caccgtggct 30000
tcgtacgtgc ctgcgttata ggtgcgcgc cgcacgtaca cgttgcgcgc cgcgtccacc 30060
ttgaccggcg tggccactg ctgcgtccgc gcccgcgcgc ggtacgttga aacgcgcatac 30120
accgggtcga tcaccagcgt ggggtgtctg tcgtagtcgg cgatcagcgg cccgcgcgc 30180
gcgcgcgcgc cggccacgaa gagctcgaag cgcgcgcgc cggcgcgcgc cgaatgcgc 30240
acctgttgcg aggctaccgg cgcgtgtgc gtaaaatttctt cctggccac ggcgcgcgc 30300
agggttgcgcgc cgcgcgcgc gtagagcgttccgcg tggccgcgcg agaggttcgg gcggatctgg 30360
cgccgcgcgc cgcgcgcgc gaccacgaag tcgtattcga cgcgcgcgc cttgcgcgtac 30420
acggtgagggt ccacgcgcgc atacagggtcc tttagcgcaca cgcgcgcgc gtcggcgc 30480
ttttcgccgc ggcggcgcc cgtgggtccgc ctgtatagt ggctgagggt ggcggcgcc 30540
tcctcgccct cgtatcggcgtt atcggaaagcc cccgcgaagc gcacgcgcgc cagcggaggcg 30600
ctggcagctt tcctggagcgt tggctcgaag gcaatgcgcgc cgcgcgcgc ggcgcgcgc 30660
ccggcgcgcgc cgcgcgcgc gtagagcgttccgcg tccggccgc actggccctc gttgcgcgc 30720
aacgtatgg cgcgcgcgcctt gagggcatac gggacgcgcgc catggacggc ggcggagggt 30780
gcgaggccgc cgcgcgcgc caccggccagg ggcggaggcg cccggccgcgc gccgcgcgc 30840
ctgcgggtca tggggcgttgc acctttctgt ttttatggct gttgttgc tttgcggat 30900
ccacagctcg gatccagctg cgcataactac cagcgcgcgc gttgttttc gtgacccagt 30960
cccgagctcg ggccccggcg acaacggcg catcgcgcgc ctttgcgcgc gcacgcgcgc 31020
cctcgatcca gtcgcccgtc cggccgcgc cggcgggggc atcgcgcgc ggcgcgcgc 31080
ccgaggccac caccggatgc cgcaggccgc cgtactcgac cagttcagg ttgcgcgcgc 31140
cgcgatggaa cgggttgcgc cgcagcgttgc ccaccgcgc gtcgagtgcc agcgcgcgc 31200
gtttctccgg gtactgcgcg atgggcacca tgcgtgcac ttccgcggcg aacggccca 31260
gtcgccgcgt gcacaggccg aggaacaccc agtcgcgcgc gcgatgcgc ggcgcgcgc 31320
cgggcgcgcgc caacttcagg tcttcccat gtcgttgcgc cccgcgcgc cccacgcgcgc 31380
gacgcgcgc accccggcg cgggtcgcgc ggcggccca acgctccgc tcgcgcgcgt 31440
tcgggatcaca cgcacgttcc ttgcgcgcgc ggcgcgcgc ctcgcgcgc ggtgcgggtgg 31500
acaccacgcg gcggtcgcac aacgcgcgcgc cgcgcgcgc gtcgagtgcc atgtccgggt 31560
agatcgtcgc cgcatacggc ttgcggcg cgcgcgcgttgc ggcgcgcgc tccaggccca 31620
ggacccatcg cgcgcgcgc tggcgccgc gacccatcg cgcgcgcgc tggtagttgt 31680
ggaagaagtt gtgcgcgcgc acggcgttgc cgtcgaggcg ctgcactcg acgcgcgttgg 31740
gcgcgcgcgc cgtggcgtgc tcgcgcgcgc acgcgcgcgc cgcgcgcgc ggcgcgcgc 31800
gcgcgcgcgc cggcgtggcg cgcgcgcgc acgcgcgcgc cgcgcgcgc ggcgcgcgc 31860
cgcagagccg aaccgcgtgc gtgtcggttgc cgcgcgcgc gacgcgcgc acggcccg 31920
agccctcgcc gcccgcgcgc aacgcgcgcgc ggttagtgcgg atcgcgcgc gaccccg 31980
gccagcgcgcgc cgcacgttccac gccttcgcg aaggaaatgac gagcttcgc accggaaacga 32040
cgccatcgcc ggcgcgcgc ggcgcgcgc ggcgcgcgc ggcgcgcgc ggcgcgcgc 32100
ggatgtcgaa gcccgcgcgc cgcagccgc ggcacagggttgc cgcacatcgcc aacggcccg 32160
cgcgctcgat ctgcgtgcgc ggcgcgcgc gacgcgcgc acgcgcgcgc 32220

gacggcccgc cgcgacgctc acctcgcgcg gcgatgcgta gagtgccgcg aggcgcctc 32280
cgccctgcgc ggcggcgga gcgcccaga cggccacgg cccgcccggc gcgattttcc 32340
agccggaaac gcgacacccg cggggccca gcaggtcggg cgcgaggca cccgtggagg 32400
ggcctgcgat gccctgcgat aggcgctoga gccagccggc gcggaaatcg ctgcaacgcg 32460
catcgacat cgcgatccac agcgtcgccg cgttgcggat ggccgcgcg aacgcggagc 32520
cggttgcgc gcccacccgga accaggtgcg tgcgcacgc gcgcccggcc agtcaccgg 32580
gccagcggtt cggccctt cgcgtcgaga tcgacgtaca cgtggccgat acgttcgcgg 32640
cagccctca cgagcgctc gatgcacgc accagccgat cccgcgcgtg cgtcgctca 32700
atgaaggcgc aaatcggtcc ggccccgatc gcgcgcgc accatgcggg cgcatcgac 32760
gcagctggca ggaacgcggc ctgcgtgcgc gtgcggcgga agtgcctccgc gaccacgcgc 32820
tggtggcgat cgcgcgcgc gtcgcttcc tgcgcggcga tggtcaggatc gaggcgatgc 32880
acgagcagat cgcgactgc caccacggcg ctgcacggc ctcctccgc gaggcgacgc 32940
gcgagatcgg tgacgcccgc ccagccgcgc gcgcgcagcc gcgcacgc cagcaacgc 33000
gcgcggcgca gcaccgcgag cccggcgagc ggctgggtgc gcgcgaactc cgcgcgcgaa 33060
gtggggacga acgcccgc gtagcggtcg ccgtgcgcgc tcaggccatc actgcagcc 33120
tagagaagct cggcggtcg gttcggtcg accgcaccccg ccaccgcgt cggcgccca 33180
tcggccagcc ggtcgccctc ctccaccaggc acgaccatt cgccttcgc gtccgcgatc 33240
gcgcgggtga cgcccccgcg cgcgcgtcg ctgcgtcgatc gtcgaggac cacggcttcc 33300
agggcgatcg cggcagtgc ctggccggc aaggcctcgatc ggcgcattca gtggcgttag 33360
gcacgcgcgc ccagcgccgt ggcgcgcgg cggcggtggaa agccgaggca cggcgccggc 33420
aacaacggat cgatcggtgc gggcacagca tccaccgcgc catcgctc cggcgccagc 33480
tcgcgataga gcgcaggta ctggccgc ttcgcctcgatc tgcgttcgg ggggtggtag 33540
ttcaggttct ccgcgcgcg atcgacgcgc gtcggtcgatc cggcgcaggcg caggaggagg 33600
gcacgcatcg cgccggatc gttgtcgcc acgaggaagc cggatcgcc atcgacgc 33660
cgctcgaa tggcgccatc ggcgcgtgccc gccaccggga tgccgaggcg ctgtgcctcg 33720
ctcaacacca ggcagaaggt ctccatc gtcgaggcgca gcaggaccatc gtcgatgcgc 33780
cgaaggatgc gcggcaggatc cgtcgccgc taggcgcgc gtcgatgtat gcccgcgc 33840
ttcgcgcgcg ctgcgagcgc gagatcgccgc ggaccccgatc ctcgatcgatc gatcgctgc 33900
ccgcgcgcgc ggcgcgcgc ctcgatgtat aggtgcgcgc ctggcgccgc gccgaaggcg 33960
ccgaggaagg ccacgcgcag caccggccgc gttcgccggc gcggaaagcag cggcaggatcg 34020
gcaatgcgt gcccgtatc ggcgcgttgc gttcgccgc tgcgcgcgc gcccgtgc 34080
atgcgcatcg ccgcgtactc ggacggcgag aggtacatcg cggccgcgc caccgcgc 34140
gacgcgcgcg cgaagcgca ttcgaggtag cgggtcactcg acgcggcgcc cgtcgccccc 34200
acgatgcct cgctgcggcc ggcaggccaa tccacgcatc cgcgcgcgc gccgcgcgc 34260
gcggcgccgc cgcacggcgatc gtcgtcgccgc cgggtcatca tggtgttagtc ggcgcacagg 34320
tacgagaagt cgtgcgcgt gatgaccacg cgtgcccccg actcacgcgc gatgcgcgc 34380
aggcgcaggc tggtccaggcc caccacgc tggaaatgcgc gacgtcgatc gccccggac 34440
agcagttcgc ggaaggccgc ctccacgcgc ggcgtcgccaa cccgcgcctc gtgcgcgc 34500
acgcggatgc cggcgatc ggcgcgtggg ttcaggcgatc cggatcgatc gtcgtcgcc 34560
aggcgctcg tggtaagtc gttccacgcgc cggggccct cggcaggcgc cagcgatcg 34620
gcgcacgttc tgcgcgcgc gttcaccaggc ggcgcgtgt gtcctcaat gcccgcgc 34680
gagtcgaggc gatgcgcgc ctcgacgcgc cggccatccgc ctcgcgtcgatc 34740
gcgcggcgatc tgacgcgcgc ggcggcgatc cggatcgatc ctcgcgtcgatc 34800
acgcgcgcgc cgttaggtgg ccagcgccgc tcgaggcgatc cgggttggc ctcgcgcgc 34860
gcgcgcgcgc cggccacccc gccaacgcag gcctcgcccg cgtggatcgatc gtagacgtcg 34920
tcgcgcgcgc cgtatcgccatc gtcgcgcgc tgcgcgcgc tgcgcgcgc 34980
ccgtagccgc ggcatacgatc cggatcgatc gggccgcatt gcacccacac ctcgcgcgc 35040
accacgcgcgc accacgcgcgc cgcgcgcgc gggccgcgc 35100

gcgacgcagg cggagatggc atcgacgtcc ggctcgccgt cggggccgctc gaagatatacg 35160
gcgacgcccgg gcaggctgag gatcgtcgcg cgggtggaga gcggggcacac gatgccgacc 35220
cgcgcatcgc tgtcgccggca acgcagcagg ccctcgagcc aaccggggcc gacctgcgt 35280
tccgaattga gcaccacgaa gtcgcccggc ctggccggca ggcccgcagc gaccgcgcct 35340
gcaagccga ggttcttcgc attggtgacg acacgagtgc cggcacgcg cgcgcgaag 35400
ctgcccagca acccggcgag gcgccggatcg gtgctcgctg catcgacgag gacgatgccc 35460
tgcgcggcgc ccgtgtggcg ggccagcgc gcgaggcgc gttcgggtgc gtcgtggccg 35520
ttgaagaccc ggacgacgc tagcgccggc gccatccgc cggcgcttcg agctgggcca 35580
atcgcccaag caggcgctgg tgctcctgcg agagtgcctc gagctgcctc tgcgaggcga 35640
cgtaggcgc tccatgcgt cccaggtgcg tgcgccctc ctcgagctgc cgcagcgaact 35700
gctcgtggcg accggagacg tcgtggtcgc ggtcggagag cagcgagaag cgcggggccc 35760
ccgcccgtcac cgccctcgcc tcgcccaca gggcgaggaa gtacatcgca tcggcgcacgc 35820
ccgcccggg ttccgcagcc tccggccggc aagcttgcag caactgcacg gcgtcgggca 35880
ccgcgtgcag cggccagatc gccgaatacg catcgacgcg ttgcccgaac agggccacct 35940
ccgggaaccc cgccttcgc acggcgagga attcgctc gtagagctcg cgcacgtat 36000
gcgggttgcg gtagtcgcgc ttgtcccgagt acacctcgag gttcggcgtc gagaccagca 36060
gcaggccgccc gggcgccagc acgcgcctcg ctgcgtcgag caggcgctcg ggatcgggaa 36120
tgtgctcgcg cgtctcgaag gagacgagga gatcgatgct tgctgcgcacgc 36180
actcgacgcg gccctcgacg tactgcaggt tctgcgcggc accatacgcc cggcgcgcct 36240
gcfgcgttgt ctcggcgccg acatccgcgc ccaccacgga ctgcgcgcg gtcgcccagca 36300
gcfgcgcggcc gtagccctcc cgcaggcga tgctcaggac gcggcagcct cggcgcagcg 36360
gcaacgcaaa gtggtagcgg tgccagtgct cgtaccagat ctgcgcctcg aagccggct 36420
ggaaacgttc gtgtccatc gggagagtt aggtggatgt gcaaccgcctc tccggagagg 36480
gcgggttgcg tgatcttatt cgagatcgcc tgctacggaa gcaccggcaa cggctgcggc 36540
ggcgcaccc accatccggt gcccgcggc acaggttcg cgtgcgcgg cgtggtgagg 36600
ccgtccatca ggcgcaccgt gatcgccca tcgacattct tgaacaccat gtcgagctt 36660
ccgtcgccat cgaagtcgag cagttgcgtg acgggtccagc cgcgcgcgc cggcagcact 36720
tccgcgcgc tgaggatggc ggtgcgcgtt atcaggcgca cgtgcgcacg gccgtcggt 36780
tggcggaaaca cgatgtcacc gcgcgcgtcg cgggtcatgt cggccacgtg gtcacacgt 36840
cagccgggtgc cgcgcgcgag gagctggaa cccgcgcgcg acgcgggtcc gttcatgt 36900
aagagatgcg cgcggccatc ggtgtggcgg aagaacatgt cgccttgcc gtcgcccgtc 36960
acgtcgccca ggtgcgtcac cgtccagccg ctggcgccgg agaggaagcc cgcgcgcgc 37020
gtgatggtgg tgccgttcat caggtatgt tagccgcggc cgtcagcgtg gatgaagacg 37080
atgtcgtccc tgccgtcgcc gttgagggtcc cccgtggcca cgcacttcca gcccgtcgcc 37140
ggccccagca gctgctggct gccgatgt gccgtccat ccatcagcca caggtgcgag 37200
cgccgcgtcg cgttgcgcag caggagatcc gccttgcgt cgcgcgttcat gtcggcggt 37260
cggtcgatgc tccagccgag gcccgcgcct agcagctcct tgccgcggcgt caccgtgagg 37320
ccgttcatcg tgcacacgtc cacgcgcggc tcgggtgtggt gggatccctc agagtcgacc 37380
tgcaggcatg caagcttgcg tattctatag tctcacctaa atagcttggc gtaatcatgg 37440
tcatagctgt ttccgtgtg aaattgttat ccgctcacaat ttcacacaac atacgagccg 37500
gaagcat 37507

<210> 3
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Sense primer.

<400> 3
ggscckcss tsdcstsga yacsgc

26

<210> 4
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Antisense
primer.

<400> 4
gcbbsryyt cdatsggRTC scc

23

<210> 5
<211> 773
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: OriT derived
from plasmid RP4.

<400> 5
gatctgtat gtacttcacc agctccgcga agtcgctctt cttgattgga ggcgcattgggg 60
acgtgtttgg caatcacgcg caccgggggg ccgttttagc ggctaaaaaa gtcatggc 120
tgcctcggg cgaccacgc ccatcatgac cttgccaagc tcgtcctgct tctttcgat 180
cttcgcgcg aggccgagga tcgtggcata accgaaccgc gccgtgcgcg ggtcgctcggt 240
gagccaggt ttcagcaggc cgcccaaggc gcccaggctg ccattgtgc gggccagctc 300
gcggacgtgc tcatagttcca cgacgcccgt gattttgtag ccctggccga cggccagcag 360
gtaggccgac aggctcatgc cggccgcgc cgcctttcc tcaatcgctc ttcttcgat 420
tggaaaggcag tacaccttga taggtggct gcccattctg gttggcttgg tttcatcagc 480
catccgcctt ccctcatctg ttacgcggc ggtagccggc cagcctcgca gagcaggatt 540
cccggttggc accgcccggt gcgaataagg gacagtgaag aaggaacacc cgctcgccgg 600
tgggcctact tcacctatcc tgccggctg acgcccgttgg atacaccaag gaaagtctac 660
acgaaccctt tggcaaaatc ctgtatatcg tgcgaaaaag gatggatata ccgaaaaaat 720
cgctataatg accccgaagc agggttatgc agcggaaaaag atccgtcgga tct 773

<210> 6
<211> 2012
<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Psi C31
integrase.

<400> 6

agatctcccg tactgacgga cacaccgaag ccccgccgc aaccctcagc ggatgccccg 60
gggcttcacg tttcccgagg tcagaagcg tttcgggag tagtccccca actggggtaa 120
ccttgagtt ctctcagttg gggcgtagg gtcggcaca tgacacaagg gttgtgacc 180
ggggtggaca cgtacgcggg tgcttacgac cgtcagtcgc gcgagcgcga gaattcgc 240
gcagcaagcc cagcgacaca gcgtagcgc aacgaagaca aggccgcga cttcagcgc 300
gaagtcgagc gcgacggggg cgggttcagg ttctcgccc atttcagcga agcgcgggc 360
acgtcggcgt tcggacggc ggagcgcgg gaggtaac gcatctgaa cgaatgcgc 420
gcggggcggc tcaacatgtat cattgtctat gacgtgtcgc gcttctcgcg cctgaagg 480
atggacgcga ttccgattgt ctcttgcatttgc gctgtgacat tgtttccact 540
caggaaggcg tcttcggca gggaaacgtc atggacctga ttccactgat tatgcggc 600
gacgcgtcgc acaaagaatc ttgcgtgaag tcggcgaaga ttctcgacac gaagaac 660
cagcgcgaat tggcgccgtt cgtcgccggg aaggcgcctt acggttcga gcttgcgg 720
gagacgaagg agatcacgcg caacggccga atggtaatg tcgtcatcaa caagcttgcg 780
caactcgacca ctcccttac cggacccttc gagttcgagc ccgacgtaat ccgggtgg 840
tggcgtgaga tcaagacgc caaacaccc ttccatcgcc cggcgcgtca agccgcatt 900
caccgggca gcatcacggg gctttgttaag cgcatggacg ctgacgcgt gcccgcgg 960
ggcgagacga ttggaaagaa gaccgcttca agcgcctggg acccgcaac ctttatgcg 1020
atccctcggg acccgcttat tgcggcttc gccgctgagg tgcgttacaa gaagaagccg 1080
gacggcacgc cgaccacgaa gattgagggt taccgcattt acgcgcaccc gatcagc 1140
cgccggcgtcg agcttgattt cggaccgtatc atcgagcccg ctgagttgtt tgactt 1200
gcgtgggttgg acggcagggg ggcggcaag gggctttccc gggggcaagc cattctgtcc 1260
gccatggaca agctgtactg cgagtgtggc gccgtcatga ttcaagcg cggggaaagaa 1320
tcgatcaagg acttttaccg ctgcgtcgc cggaaagggtgg tcgaccgcgc cgcacctgg 1380
cagcacgaag gcacgtgcaaa cgtcagcatg gcccactcg acaagttcg tgcggaaacgc 1440
atcttcaaca agatcaggca cgccgaaggc gacgaagaga cgttgcgc tctgtggaa 1500
gccgcccac gcttcggcaa gctcaactgag ggcctgaga agacggcga acgggcgaac 1560
tttgcgttgcgg agcgcgcgca cggccctgaac gcccgttgaag agctgtacga agaccgcgc 1620
gcaggcgcgt acgacggacc cttggcagg aacgacttcc ggaagcaaca ggcagcgcgt 1680
acgctccgcg acgaaggggc ggaagagcgg cttgcgcgaac ttgaagccgc cgaagccccg 1740
aagcttcccc ttgaccaatg gttcccgaa gacgcccacg ctgacccgac cggccctaag 1800
tcgtgggtgg ggcgcgcgtc agtagacgc aacgcgcgtt tcgtgggtt cttcgtagac 1860
aagatcggtt tcacgaagtc gactacggc agggggcagg gaacgcgcacat cgagaagcgc 1920
gcttcgatca cgtggcgaa gccgcgcacc gacgacgcacg aagacgcgc ccaggacggc 1980
acggaagacg tagcggcgta gcgagacacc cg 2012

<210> 7

<211> 84

<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Left arm of
pPAOI6 transposon.

<400> 7
ctgtcttta tacacatctc aaccatcatc gatgaatttt ctcgggtgtt ctcgcatttt 60
84
ggctcgaatt cgagctcggt accc

<210> 8
<211> 94
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Rigth arm of
pPAOI6 transposon.

<400> 8
gatcctctag agtcgacctg caggcatgca agcttgccaa cgactacgca ctagccaaca 60
94
agagcttcag ggttgagatg tgtataagag acag

<210> 9
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer AmF.

<400> 9
ccctaagatc tggttcatgt gcagctccat c 31

<210> 10
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer AmR.

<400> 10
tagtaccgg ggatccaacg tcatctcggtt ctcc 34

<210> 11

<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer OriTF.

<400> 11
gccccatgtc tgtgtatgtac ttcaccagct cc

32

<210> 12
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer OriTR.

<400> 12
tagtacccgg ggcgttgcacg gatctttcc gctgcac

37

<210> 13
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer Fint.

<400> 13
aacaaggatc tcccgtaactg acggacacac cg

32

<210> 14
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer PJ.

<400> 14
cggtgtctc gcatcgccgc t

21

<210> 15

<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer
amyE-BamHI

<400> 15
atcgcaggat cctgaggact ctgcgaacccg

30

<210> 16
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer
amyE-EcoRI

<400> 16
cgactgaatt cagatctagc gtgtaaattc cgtctgc

37

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 03/07765

A. CLASSIFICATION OF SUBJECT MATTER
 IPC 7 C12N15/10 C07K14/00 C07H21/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K C07H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EMBL, EPO-Internal, BIOSIS, MEDLINE, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>WO 00 78977 A (AUGUST PAUL ;MACNEIL IAN (US); GROSSMAN TRUDY (US); AVENTIS PHARMA) 28 December 2000 (2000-12-28) page 5, line 1 - line 11 page 11, line 17 - line 19 page 14, line 13 - line 17 claim 21</p> <p>---</p> <p>-/-</p>	1-5,7,8, 18,19

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

& document member of the same patent family

Date of the actual completion of the international search

7 January 2004

Date of mailing of the international search report

23/01/2004

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
 Fax: (+31-70) 340-3016

Authorized officer

Voigt, H

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 03/07765

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MEJIA J E ET AL: "RETROFITTING VECTORS FOR ESCHERICHIA COLI-BASED ARTIFICIAL CHROMOSOMES (PACS AND BACS) WITH MARKERS FOR TRANSFECTION STUDIES" GENOME RESEARCH, COLD SPRING HARBOR LABORATORY PRESS, US, vol. 7, no. 2, February 1997 (1997-02), pages 179-186, XP001106025 ISSN: 1088-9051 page 179, column 2, line 14 - line 21 page 181, column 2, paragraph 2 page 183, column 2, paragraph 1 page 184, column 2, line 20 - line 27 ---	1,2,4,5
X	KIM SU YOUNG ET AL: "Modification of bacterial artificial chromosome clones using Cre recombinase: Introduction of selectable markers for expression in eukaryotic cells." GENOME RESEARCH, vol. 8, no. 4, April 1998 (1998-04), pages 404-412, XP002232358 ISSN: 1088-9051 page 409, column 1, line 3 - line 8 ---	1,4-6
X	DATABASE EMBL 'Online' 23 November 2000 (2000-11-23) "Arabidopsis thaliana chromosome 1" retrieved from EBI Database accession no. AC084820 XP002232359 abstract ---	22
Y	EP 0 176 321 A (LILLY CO ELI) 2 April 1986 (1986-04-02) page 3, line 23 - line 28 ---	1-21
Y	WO 98 17811 A (CHROMAXOME CORP) 30 April 1998 (1998-04-30) page 7, line 28 - line 33 page 9, line 8 - line 11 ---	1-21
A	WO 01 40497 A (AVENTIS PHARMA SA ; PERNODET JEAN LUC (FR); COURTOIS SOPHIE (FR); J) 7 June 2001 (2001-06-07) the whole document ---	1-21
		-/-

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 03/07765

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>RONDON M R ET AL: "CLONING THE SOIL METAGENOME: A STRATEGY FOR ACCESSING THE GENETIC AND FUNCTIONAL DIVERSITY OF UNCULTURED MICROORGANISMS"</p> <p>APPLIED AND ENVIRONMENTAL MICROBIOLOGY, WASHINGTON, DC, US, vol. 66, no. 6, June 2000 (2000-06), pages 2541-2547, XP000946829</p> <p>ISSN: 0099-2240</p> <p>the whole document</p> <p>-----</p>	1-21

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-21

Method of analysing a library of polynucleotides comprising the selection of cloning vectors, modification of these vectors and analysis the polynucleotides contained in vectors upon transfer in host cell.

2. Claims: 22-23

Polynucleotide sequence comprising all or part of SEQ ID NO:1 or 2. An oligonucleotide comprising SEQ ID NO:3 or 4.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP 03/07765

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.: because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 03/07765

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 0078977	A 28-12-2000	AU 5496600 A BR 0011765 A CA 2375821 A1 EP 1196613 A1 JP 2003503029 T NO 20016153 A WO 0078977 A1 ZA 200109441 A		09-01-2001 05-03-2002 28-12-2000 17-04-2002 28-01-2003 15-02-2002 28-12-2000 17-02-2003
EP 0176321	A 02-04-1986	AU 592956 B2 AU 4787785 A CA 1238591 A1 DE 3572510 D1 DK 429685 A EP 0176321 A1 HU 38676 A2 IE 58265 B1 JP 8015432 B JP 61085191 A US 4921801 A		01-02-1990 10-04-1986 28-06-1988 28-09-1989 28-03-1986 02-04-1986 30-06-1986 25-08-1993 21-02-1996 30-04-1986 01-05-1990
WO 9817811	A 30-04-1998	US 5783431 A AU 744960 B2 AU 5163298 A CA 2269123 A1 EP 0951557 A1 JP 2001502547 T KR 2000052795 A WO 9817811 A1 US 6242211 B1 ZA 9709557 A		21-07-1998 07-03-2002 15-05-1998 30-04-1998 27-10-1999 27-02-2001 25-08-2000 30-04-1998 05-06-2001 12-05-1998
WO 0140497	A 07-06-2001	FR 2801609 A1 AU 2179101 A BR 0015993 A CA 2393041 A1 EP 1268764 A2 WO 0140497 A2 JP 2003520578 T NO 20022532 A		01-06-2001 12-06-2001 06-08-2002 07-06-2001 02-01-2003 07-06-2001 08-07-2003 26-06-2002

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.